Investigation of Whole Genome Phylogeny and Genetic Analysis of Cyclic Lipopeptide Biosynthesis Genes in Antagonistic Strains of *Bacillus*

by

Emma Groulx

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Science

in

Biology

To

Carleton University
Ottawa, Ontario

Acknowledgements

I would first like to thank Dr. Myron Smith and Dr. Tyler J. Avis for being my co-supervisors and providing me with support and guidance throughout my graduate studies. I greatly appreciate the opportunity to continue my education and gain more research experience.

Special thanks go out to Dr. Alex Wong and Dr. Nicolas Rodrigue for their guidance and assistance in data analysis and interpretation of results.

I would like to thank Prabhjeet Basra for assistance with sequencing protocols and assessment of sequencing data.

I would also like to thank Thomas Bujaki for technical assistance and for emotional support over the past two years.

Lastly, I would like to thank all my family and friends who supported me throughout my graduate years.

Abstract

Disease suppressive composts and soil were found to contain bacteria belonging to the genus *Bacillus* that were antagonistic to plant pathogens. It is known that these species of bacteria produce antimicrobial compounds called cyclic lipopeptides. These bacteria produce three families of lipopeptides (surfactins, fengycins and iturins), however not all the isolates produce all of the lipopeptides. The genetics behind the production of these lipopeptides is not as well known. The objective of this study was to confirm the identity of the bacterial isolates using whole genome phylogeny and identify the *Bacillus* cyclic lipopeptide biosynthesis genes. Fifteen isolates of *Bacillus subtilis*, two isolates of *B. megaterium* and one isolate each of *B. badius* and *Brevibacillus borstelensis* were sequenced using an Illumina Miseq system and phylogenetic analysis was performed using protein-coding alignments. The cyclic lipopeptide genes were identified using nucleotide and protein blasts. It was shown that the identity of the bacterial isolates resembled the identity assigned using 16S rRNA sequences, with the exception of *B. subtilis* isolates CU12 and M9-3. It was also determined that the *B. subtilis* isolates and the *B. borstelensis* isolate possessed the gene operons for the cyclic lipopeptides whereas *B. megaterium* and *B. badius* isolates did not.

Key Words: *Bacillus*, biocontrol, *Brevibacillus*, cyclic lipopeptides, phylogenetics.

Table of contents

List of tables

List of figures

Chapter 1: Introduction

1.1. Food security

With the population reaching nine billion by the middle of this century, the global demand for food will increase. This higher demand leads to an increase in the need for food products, putting pressure on the food supply system (Godfray et al., 2010). Food production systems need to become fully sustainable in order to reduce the amount of land and resources utilized as well as the amount of loss due to pests (Godfray et al., 2010). Pests, which are harmful organisms such as insects, rodents, weeds, bacteria and fungi, possess the ability to drastically reduce crop production and compromise food security (Oerke, 2005). Pests are estimated to cause between 27% and 42% losses in major crops around the world, and these losses would increase to between 43% and 83% without crop protection (Glare et al., 2012). Along with weeds and insects, fungal plant pathogens are a major cause of crop loss globally (Oerke, 2005). Synthetic chemical fungicides are among the most frequent methods used to reduce disease and the amount of loss due to fungal pathogens.

1.2. Fungicides

The use of chemical compounds to control fungal pathogens began more than a century ago with the use of simple fungicides that act as surface protectants. These first-generation fungicides are plant protectants that consist of noxious chemicals such as organic mercury, sulphur and copper, and are not capable of penetrating plant tissue. Second-generation fungicides are considered systemic fungicides, and these compounds possess the ability to penetrate the plant tissue and provide a higher level of protection. However, these compounds act on a small number

of target sites, which increases the possibility of the fungal pathogen developing resistance. Third-generation fungicides consist of systemic compounds that indirectly enhance plant defense mechanisms to help the plant fight off disease (Russell, 2005).

As described above, the overall disadvantages with first- and second-generation fungicides, protectant or systemic, are the potential adverse effects to human health and the environment and risks of the pathogens developing resistance over time. This has led to the development of alternative methods of controlling pathogens, including cultural methods such as crop rotation, genetic methods i.e., use of genetically resistant crops through breeding or genetic engineering, and biological methods such as using microbial biological control agents (Glare et al., 2012).

1.3. Biological control

Biological control is the use of living organisms to control pests such as weeds, insects, and plant diseases. These organisms include nematodes, insects, and microorganisms (Ongena & Jacques, 2008). Of the commercially-available biopesticide products on the market, about 30% are microorganism-based.

These microbial agents typically possess one or more of the main biocontrol modes of action: (i) competition for nutrients or space, (ii) induced plant resistance, (iii) parasitism, and/or (iv) antibiosis (Sharma, Singh, & Singh, 2009). In the particular case of antibiosis, inhibition of pathogens occurs through the production of antimicrobial compounds, such as toxins or antibiotics, by an antagonist. These metabolites are compounds produced by organisms that are not necessary for reproduction and growth, and are the result of specialized metabolism within an organism (Vining, 1990). These antimicrobial secondary metabolites inhibit primary processes that are

essential to other microorganisms. It has been hypothesized that microorganisms regulate the production of these compounds and optimize them to modify the metabolism, inhibit the growth, or kill potential competitors to provide themselves with a survival advantage (Gonzalez et al., 2011).

1.4. Composts as sources of microbial control agents

Composts, which contain decomposed organic matter such as animal waste and plant debris, are used in agriculture to provide nutrients to the soil and are sources of beneficial microorganisms. Some of the microorganisms present in composts have been shown to produce antimicrobial compounds (Niisawa et al., 2008). *Bacillus* spp. are among the most frequently reported antagonistic bacteria in composts (Kavroulakis et al., 2010). In a recent study in our laboratory, Mohamed et al. (2017) reported that composts contained numerous bacteria from this and related genera such as *Brevibacillus*.

1.5. *Bacillus* and related bacteria

1.5.1. *Bacillus*

Bacillus is a genus of Gram-positive bacteria that is found in many different environments, such as water, soil and in extreme environments such as high pH and temperature. These rhizobacteria are rod-shaped and possess the ability to form endospores, which allow them to survive environmental stress and harsh conditions. *Bacillus* bacteria are model organisms that are used quite frequently in molecular biology and genetics (Harwood, 1992).

Bacillus subtilis (Ehrenb.) Cohn is a species of *Bacillus* that has provided valuable information on the physiology, genetics, and biochemistry of Gram-positive bacteria, especially on metabolism and sporulation (Harwood, 1992). Hundreds of strains of *B. subtilis* have been isolated with the ability to produce a vast variety of antimicrobials, including peptide and non-peptide compounds (Moszer, Jones, Moreira, Fabry, & Danchin, 2002). The production of these antimicrobials has provided some *B. subtilis* strains with the ability to control various plant diseases caused by fungal pathogens. Different strains of the bacterium have been shown to inhibit the growth of *Fusarium oxysporum* on pepper plants (Yu, Ai, Xin, & Zhou, 2011) and suppress *Alternaria solani* as well as *Rhizoctonia solani* (Grover, Nain, Singh, & Saxena, 2010). In addition, *B. subtilis* has been shown to reduce the incidence of sheath blight of rice caused by *Sclerotinia sclerotiorum* (Yang, Wang, Wang, Chen, & Zhou, 2009).

1.5.2. *Brevibacillus*

A reclassification of strains in the group *Bacillus brevis* lead to the formation of the genus *Brevibacillus* in 1996. This genus is made up of 20 species (Panda, 2014), and these bacteria reside in a variety of diverse environments, such as water, dust and inside the guts of various animals and insects (Nicholson, 2002). Many members of this genus are of biological significance and can act as potential biocontrol agents. *Brevibacillus* spp. possess the ability to produce antimicrobial compounds, which can control insects from the Lepidoptera and Coleoptera orders (Oliveira et al., 2004), as well as plant diseases (Sunita et al., 2010). An analysis of the *B. brevis* X23 genome identified gene clusters related to polyketide synthetases and nonribosomal peptide synthetases (Zhou et al., 2012), which are involved in antimicrobial production.

1.6. Cyclic lipopeptides

Cyclic lipopeptides (CLPs) are a group of antimicrobial compounds that are made of two main components: a cyclic peptide ring and a fatty acid chain (Raaijmakers, De Bruijn, Nybroe, & Ongena, 2010). These compounds are produced via nonribosomal pathways using large multifunctional enzymes (Cane, Walsh, & Khosla, 1998). In the specific case of *Bacillus* spp., cyclic lipopeptides are generally classified into three different families: fengycins, surfactins and iturins (Figure 1), and these families differ structurally in terms of the amino acids that make up their peptide rings, and the length of their fatty acyl chain (Ongena & Jacques, 2008).

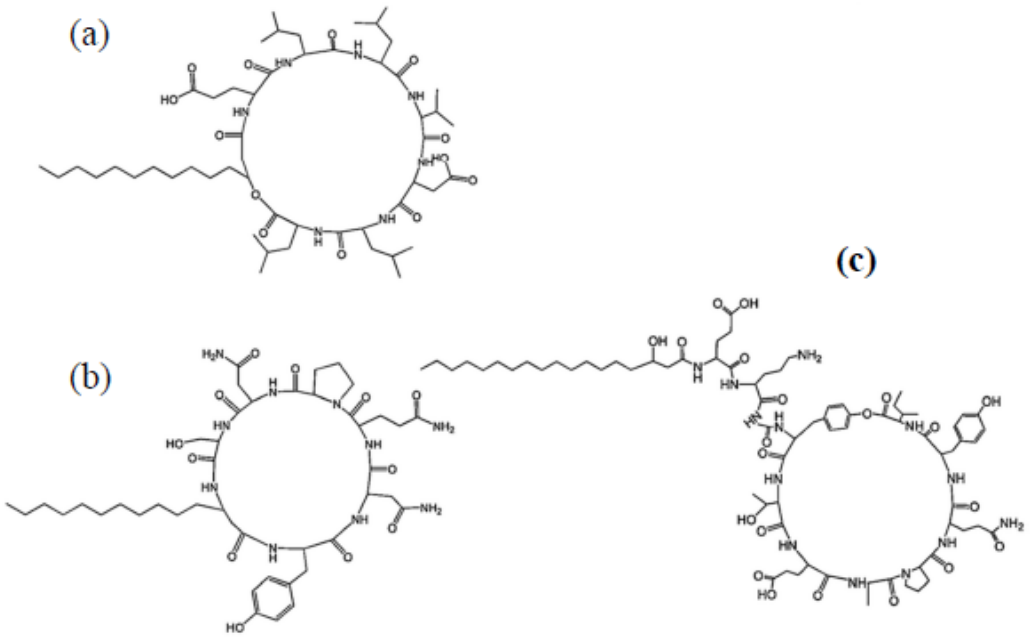

Figure 1: Structural representation of three families of lipopeptides produced by *Bacillus* spp. (a) surfactin (b) iturin and (c) fengycin (Ongena *et al*., 2005)

1.6.1. Surfactin

Surfactins are amphiphilic compounds that are composed of β-hydroxy fatty acyl chain, varying from 13 to 15 carbons, bound to a cyclic heptapeptide structure. Surfactins possess antibacterial (Bais, Fall, & Vivanco, 2004), antiviral and hemolytic activity, but limited antifungal activity (Perez-Garcia, Romero, & de Vicente, 2011; Peypoux, Bonmatin, & Wallach, 1999). However, there has been research indicating the ability of surfactin to inhibit the germination of mold spores (Liu, Hagberg, Novitsky, Hadj-Moussa, & Avis, 2014; On et al., 2015; Ongena & Jacques, 2008)

The antimicrobial activity of surfactin comes from its ability to alter the integrity of cellular membranes. This lipopeptide directly inserts itself into the membrane and forms ion-conducting channels, which can cause the leakage of cellular contents (Carrillo, Teruel, Aranda, & Ortiz, 2003; Peypoux et al., 1999). These membrane-disrupting properties are influenced by how big the peptide ring is, how strong the peptide binds to the membrane, and how surfactin enters the membrane (Heerklotz, Wieprecht, & Seelig, 2004). In addition to this direct effect on microbial cell membranes, surfactin has been shown to induce defense responses in plants, known as induced systemic resistance (Ongena & Jacques, 2008).

1.6.2. Iturin

Members of the iturin family, which include bacillomycin, mycosubtilin and iturin A, are characterized by cyclic heptapeptide rings bound to a β-amino fatty acid chain varying in length from 14-17 carbon atoms. Iturins possess strong antifungal and hemolytic activity, but have limited antibacterial and antiviral activity (Maget-Dana and Peypoux, 1994). They affect

pathogens by interacting with the phospholipids, creating channels in the plasma membrane, which allows for the release of cellular contents. Iturin A was shown to inhibit the growth of fungal pathogens such as *Podosphera fusca*, (Romero et al., 2007), *Alternaria citri*, *Colletotrichum gloeosporioides* and *Penicillium crustosum* (Arrebola, Jacobs, & Korsten, 2010).

1.6.3. Fengycin

Fengycins are composed of a β-hydroxy fatty acid chain that is bound to a decapeptide, where eight of the amino acids form a lactone ring structure. The fatty acid chain varies in length from 14-18 carbons, and structural changes result in different homologs of fengycin: fengycin A and fengycin B (Deleu, Paquot, & Nylander, 2008). Fengycin A has D-alanine in position six of the peptide chain, whereas fengycin B has D-valine. Plipastatin A and plipastatin B are structurally nearly identical to their fengycin counterparts, which the exception of the configuration of the tyrosine in position three of the peptide chain (Eeman, Pegado, Dufrêne, Paquot, & Deleu, 2009). Fengycin homologs have the tyrosine in the D-isomer form, whereas plipastatin homologs have the tyrosine in the L-isomer form (Tsuge, Ano, Hirai, Nakamura, & Shoda, 1999)

Fengycins are not as hemolytic as surfactins or iturins, but possess strong antifungal properties, specifically against filamentous fungi (Deleu et al., 2008). Fengycins interact with plasma membranes of pathogens, causing pore formation and leakage of cellular contents. They readily insert into the plasma membrane in an all-or-none mechanism – there is no effect on the membrane when the concentration of fengycin is low, but there is pore formation and cell death when the concentration of fengycin is high (H. Patel, Tscheka, Edwards, Karlsson, & Heerklotz, 2011). Studies have shown fengycin to be an inhibitor of *Pythium ultimum* and *Botrytis cinerea*

(Ongena et al., 2005) and bacterial strains that produce fengycin were shown to inhibit *Fusarium graminearum* and *S. sclerotiorum* (Ramarathnam et al., 2007).

1.6.4. Nonribosomal peptide synthetases of *Bacillus* cyclic lipopeptides

Nonribosomal Peptide Synthetases (NRPS) are large multimodular enzymes that produce peptides through nonribosomal pathways (Cane et al., 1998). These complexes are composed of modules, which are responsible for adding amino acids to the peptide chain, and there are four essential domains. There is an adenylation (A) domain, which is responsible for selecting and activating the amino acids in the peptide. The peptidyl carrier protein (PCP) domain is responsible for accepting the amino acids and holding onto the growing peptide chain. The condensation (C) domain catalyzes the peptide bond between the incoming amino acid and the PCP domain of adjacent modules. The last domain is the thioesterase (TE) domain, and this domain is responsible for terminating the reaction and catalyzing the release of the peptide chain (Kopp & Marahiel, 2007).

The cyclic lipopeptides produced by *B. subtilis* are synthesized using these nonribosomal synthetases and the genes responsible for the production of these peptides are organized in an operon. Fengycin and plipastatin biosynthesis is encoded by the fen/pps operon, which contains a three-module enzyme made up of five genes. The first three genes (*fenC*, *fenD*, *fenE*) code for two-module enzymes that activate the first six amino acids in the peptide. The fourth gene (*fenA*) codes for the next three amino acids, and the fifth gene (*fenB*) codes for the last amino acid. *FenC* activates glutamate and ornithine, *fenD* thiolates tyrosine and allothreonine, *fenE* integrates glutamate and valine (alanine in the case of plipastatin). *FenA* activates proline, glutamine, and

tyrosine and *fenB*, the final gene, activates isoleucine and possibly facilitates in the closing of the fengycin ring (Steller & Vater, 2000)

The surfactin operon is comprised of four biosynthetic genes; three of which are responsible for the peptide synthetases and the fourth encoding for a thioesterase protein. The first gene, *srfAA*, contains three modules and incorporates the first three amino acids into the peptide (glutamate, leucine and leucine). The second gene, *srfAB*, also contains three modules and incorporates valine, aspartate, and leucine. The third gene, *srfAC*, activates leucine and catalyzes the release and cyclization of the peptide chain. The final gene, *srfAD*, encodes for a protein that is involved in the repair of incorrect amino acids during peptide biosynthesis (Cosmina et al., 1993; Tseng et al., 2002).

The gene cluster for iturin A contains four biosynthetic genes: *ituD*, *ituA*, *ituB*, and *ituC*. *ItuD* is the first gene and it encodes for a malonyl-CoA transacylase. The second gene, *ituA*, possesses one module and incorporates the first amino acid (asparagine) into the peptide. The third gene, *ituB*, contains four modules and incorporates the next four amino acids (tyrosine, asparagine, glutamine and proline). The final gene, *ituC*, activates asparagine and serine, and is responsible for peptide cyclization (Tsuge, Akiyama, & Shoda, 2001).

1.7. Species identification

DNA-DNA hybridization was considered the "gold standard" method of identifying new bacterial species, but because this method is quite time-consuming and expensive, the use of 16S rRNA sequences is often the current tool for classifying bacteria (Janda & Abbott, 2007). The 16S gene is present in all bacteria, it is long enough for informatic purposes (approx. 1,500 bp), and its

function has remained the same over time which indicates a more accurate measure of evolution (J. B. Patel, 2001). However, closely related bacterial species cannot always be differentiated using this gene due to the extremely slow rate of evolution. Whole-genome sequences can offer valuable information to aid in the identification and classification of bacteria (Gomila, Peña, Mulet, Lalucat, & García-Valdés, 2015).

1.8. Objectives and hypotheses

1.8.1. Objectives

The objectives of this study were (i) to confirm or correct the identity of isolates antagonistic to plant pathogens using whole genome phylogeny and (ii) to identify genes implicated in the biosynthesis of *Bacillus* cyclic lipopeptides.

1.8.2. Hypotheses

(i) The identity of the bacterial isolates determined by whole genome sequencing will match the identity assigned by 16S rRNA sequencing.

(ii) Bacterial isolates will possess biosynthesis genes from different families of cyclic lipopeptides that correlate with known lipopeptide production.

Chapter 2: Materials and methods

2.1. Bacteria material

Fifteen isolates of *B. subtilis*, two isolates of *Bacillus megaterium* and one isolate each of *Bacillus badius* and *Brevibacillus borstelensis* (Table 1) were used for this project. All bacteria were isolated from agricultural composts or soil and were previously shown to possess antagonistic activities against plant pathogens (Bernier-English, Avis, Mimee, Antoun, & Tweddell, 2010; Martin-Lapierre, Mimee, & Tweddell, 2011; Wise, Falardeau, Hagberg, & Avis, 2014; Wise, Novitsky, Tsopmo, & Avis, 2012). The bacterial identification were tentatively assigned following sequencing of 16S rDNA and comparison of the sequences to curated type strains in the Ribosomal Database Project (Michigan State University) as well as phylogenetic inferences using these 16S rDNA sequences (Mohamed et al., 2017).

Table 1: Identity and source of bacteria isolated from compost material

Isolate	Source	Identity
B9-1	Bovine manure compost	*Bacillus subtilis*
B9-5	Bovine manure compost	*Bacillus subtilis*
B9-7	Bovine manure compost	*Bacillus subtilis*
B9-8	Bovine manure compost	*Bacillus subtilis*
B9-9A	Bovine manure compost	*Bacillus subtilis*
B9-9B	Bovine manure compost	*Bacillus megaterium*
B9-14	Bovine manure compost	*Bacillus subtilis*
CU12	Suppressive soil	*Bacillus subtilis*
F9-2	Forestry residue compost	*Bacillus subtilis*
F9-8	Forestry residue compost	*Bacillus subtilis*
F9-12	Forestry residue compost	*Bacillus subtilis*
M9-1B	Marine residue compost	*Bacillus megaterium*
M9-3	Marine residue compost	*Bacillus subtilis*
M9-4	Marine residue compost	*Bacillus subtilis*
M9-7	Marine residue compost	*Bacillus subtilis*
M9-9	Marine residue compost	*Bacillus subtilis*
M9-14	Marine residue compost	*Bacillus subtilis*
M9-18	Marine residue compost	*Brevibacillus borstelensis*
M9-20	Marine residue compost	*Bacillus badius*

2.2. Whole-genome sequencing and assembly

The bacterial isolates were cultured from a glycerol stock and grown overnight in tryptic soy broth (TSB, Becton Dickinson, Sparks, Maryland). Genomic DNA from these isolates was extracted using the ONE-4-ALL Genomic Mini-Preps Kit (Bio Basic Canada, Markham, Ontario) following the manufacturer's protocols for Gram-positive bacteria. After DNA purification, DNA was quantified using the Quant-iT^{TM} PicoGreen® dsDNA Assay Kit (Invitrogen, Waltham, Massachusetts). The Nextera XT kit (Illumina, San Diego, California) was used to create libraries and those libraries were quantified by qPCR (KAPA Biosystems, Wilmington, Massachusetts). The Illumina Miseq system was used for the sequencing of paired-end reads that were 300 bp in length.

The quality of the raw sequence reads was determined using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and Trimmomatic-0.32 was used to removed low quality data from the sequence reads (Bolger, Lohse, & Usadel, 2014). The program was run using the command line version with parameters that removed up to 20 bases from the start of the reads and removed bases to make the reads a specified length. Also, reads below a base-call quality threshold of 20 were removed, when scanning a 4-bp window. The parameters varied between bacterial sequence reads, as each bacterial sequence was not of the same quality and each required a different number of bases removed.

De novo assemblies were performed using VelvetOptimiser-2.2.5 to optimize the *k-mer* value for the assemblies and Velvet-1.2.10 (Zerbino & Birney, 2008). The quality of the assemblies was determined by assessing the n50 value and number of contigs produced by Velvet for each assembly. The assemblies were annotated using Rapid Annotation using Subsystem Technology (RAST) (Brettin et al., 2015).

2.3. Multiple sequence alignment and phylogenetic analysis

Multiple sequence alignments of whole genomes were performed with progressiveMauve (Darling, Mau, Blattner, & Perna, 2004). An alignment was performed using the annotated genomes of the fifteen *B. subtilis* isolates, along with one isolate of *B. badius* and two isolates of *B. megaterium* using default parameters and eighteen reference genomes (Table 2).

Table 2: Reference genomes in *Bacillus* phylogenetic tree, obtained from NCBI

Reference strain	GenBank accession number
Bacillus acidicola strain FJAT-2406	GCA_001636425.1
Bacillus amyloliquefaciens strain B15	CP014783.1
Bacillus badius strain DSM30822	GCA_000829555.1
Bacillus flexus strain Riq5	GCA_001238505.1
Bacillus koreensis strain DSM16467	GCA_001274935.1
Bacillus licheniformis strain HRBL-15TD17	CP014781.1
Bacillus megaterium strain ATCC145881	GCA_000832985.1
Bacillus methylotrophicus strain B25	LN999829.1
Bacillus mojavensis strain RRC101	GCA_000404285.1
Bacillus oleronius strain DSM9356	GCA_002009555.1
Bacillus shackletonii strain LMG18435	GCA_001420715.1
Bacillus siamensis strain SRCM10069	GCA_001662915.1
Bacillus sporothermodurans strain B4102	GCA_001587375.1
Bacillus subtilis strain 168	GCA_000789275.1
Bacillus tequilensis strain FJAT-14262a	GCA_001278955.1
Bacillus vallismortis strain B4144_201601	GCA_001587405.1
Domibacillus robinginosus strain WS4628	GCA_000966195.1
Planococcus rifietoensis strain M8	CP013659.2

Another alignment was performed with the *B. borstelensis* isolate, using default parameters and nine reference genomes (Table 3).

Table 3: Reference genomes in *Brevibacillus* phylogenetic tree, obtained from NCBI

Reference strain	GenBank accession number
Brevibacillus agri strain BAB-2500	GCA_000328345.1
Brevibacillus borstelensis strain cifa_chp40	GCA_000738785.1
Brevibacillus brevis strain NBRC 110488	GCA_001748185.1
Brevibacillus choshinensis strain DSM8552	GCA_1420695.1
Brevibacillus formosus strain DSM9885	GCA_00102775.1
Brevibacillus laterosporus strain B9	CP011074
Brevibacillus panacihumi strain W25	GCA_0005037775.1
Brevibacillus parabrevis strain CN1	GCA_001619605.1
Brevibacillus reuszeri strain DSM9887	GCA_001187725.1

Three different settings of Gblocks were used to remove positions in the protein-coding alignments that were poorly aligned (Castresana, 2000), then the Robinson-Foulds (RF) metric was calculated on the resulting trees (Robinson & Foulds, 1981). The RF calculation determined that there were no discernable differences between the trees. The alignment used to construct the phylogenetic trees had positions removed by Gblocks if 50% or more of the sequences had a gap at that position. PhyML was used to construct the phylogenetic tree with generalized time reversible (GTR) model of substitution, with gamma distribution and proportion of invariable sites estimated by maximum likelihood. The appropriate likelihood ratio (aLRT) was run, returning Chi^2-based parametric branch supports (Guindon et al., 2010).

2.4. Identification of CLP genes

The genes responsible for the biosynthesis of CLPs were identified using the NCBI basic local alignment search tool (BLAST) on a local Linux machine. Databases were first generated using the bacteria samples, and then the genes for fengycin, plipastatin, surfactin, and iturin were

identified using the traditional nucleotide blast option and the discontiguous megablast (dc-megablast) option (Altschul, Gish, Miller, Myers, & Lipman, 1990; Camacho et al., 2009).

These results were confirmed by performing a protein BLAST of the hits made by the local nucleotide blast. The nucleotide sequences of the blast hits were translated into amino acid sequences and a protein blast was performed of these sequences on the NCBI website, searching the reference proteins (refseq_protein) database and restricting the organism hits to bacterial proteins (taxid: 2).

Chapter 3: Results

3.1. General assembly statistics

The genome assemblies of the bacterial isolates were created using VelvetOptimiser-2.2.5 to optimize the *k-mer* value for the assemblies and Velvet-1.2.10. The quality of the assemblies was assessed using the n50 value and number of contigs for each assembly (Table 4). The average coverage of each base in the genome assemblies is 49 and the average read length is 221 base pairs (Table 4).

Table 4: Statistics for the *Bacillus* and *Brevibacillus* genome assemblies

Isolate	Coverage	N50 Value	Length of Reads (bp)	Number of Reads	Number of Contigs
B. subtilis B9-1	52.9	4298805	200	1,115,968	113
B. subtilis B9-5	14.7	3709091	200	311,233	100
B. subtilis B9-7	58.7	4337818	210	1,179,785	103
B. subtilis B9-8	46.7	3016940	200	986,000	109
B. subtilis B9-9A	73.5	4330637	200	1,550,676	113
B. subtilis B9-14	27.5	4348569	225	516,832	127
B. subtilis CU12	36.1	4144869	250	608,910	111
B. subtilis F9-2	59.3	4261308	250	1,000,113	275
B. subtilis F9-8	34.6	4364092	220	664,364	128
B. subtilis F9-12	38.2	4267377	250	644,630	123
B. subtilis M9-3	75.9	4113131	250	1,280,435	94
B. subtilis M9-4	48.9	4215630	200	1,030,826	163
B. subtilis M9-7	54.1	4319104	200	1,140,133	114
B. subtilis M9-9	67.5	4331858	225	1,265,544	122
B. subtilis M9-14	67.4	4308915	220	1,291,472	146
B. badius M9-20	55.4	5657035	200	1,127,589	125
B. megaterium B9-9B	41.9	5367571	220	1,095,503	299
B. megaterium M9-1B	34.7	5275229	260	769,133	293
B. borstelensis M9-18	44.2	5234430	220	1,045,057	219

3.2. Phylogenetic analysis

A multiple sequence alignment of the *Bacillus* and *Brevibacillus* isolates was generated with progressiveMauve using default parameters. Phylogenetic analyses were performed on the protein-coding sections of a multiple sequence alignment of the *Bacillus* and *Brevibacillus* isolates.

Trees were generated with the GTR model of substitution, with gamma distribution and proportion of invariable sites estimated by maximum likelihood. The approximate likelihood ratio statistic was used to determine branch support.

3.2.1. *Brevibacillus* phylogenetic relationships

Figure 2 shows the phylogenetic relationships among the *Brevibacillus* strains based on whole genome sequencing data. One isolate of *B. borstelensis*, previously identified by 16S rRNA sequences, was used with nine reference genomes obtained from GenBank. There were 2,682,577 positions in the protein coding alignment after Gblocks removed positions for which half or more of the species had a gap at that position, which was nineteen percent of the total alignment (19,970,069 positions). The clustering of isolate M9-18 with *B. borstelensis* strain cifa_chp40 is highly supported (aLRT value > 0.99), indicating that M9-18 is most closely related to *B. borstelensis*.

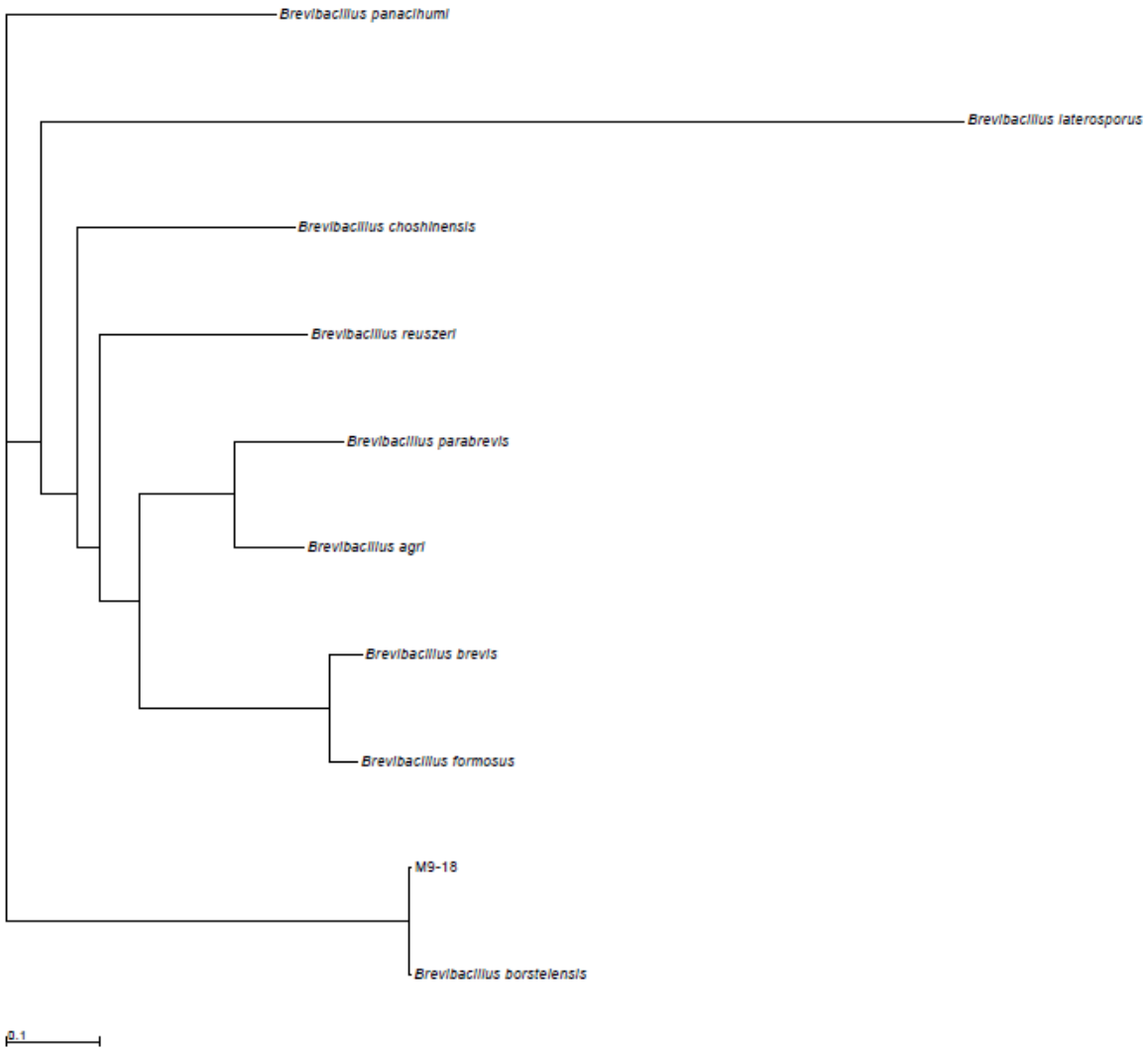

Figure 2: Maximum likelihood phylogenetic tree based on the general time reversible model of substitution of whole-genome sequences of *Brevibacillus* isolates. Branch length is represented in expected number of substitutions per site. Node support is approximate likelihood ratio; all nodes have a value of 0.99985.

3.2.2. *Bacillus* phylogenetic relationships

Figure 3 shows the phylogenetic relationships of *Bacillus* strains based on whole genome sequencing data. Fifteen previously identified isolates of *B. subtilis*, along with two isolates of *B. megaterium* and one isolate of *B. badius* were used with eighteen reference genomes obtained from GenBank. There were 713,067 positions in the protein coding alignment after Gblocks removed positions for which half or more of the species had a gap at that position, which was three percent of the total alignment (19,970,069 positions).

When looking at the section of the *Bacillus* phylogenetic tree that contains the *B. badius* and *B. megaterium* isolates, isolate M9-20 clusters with *B. badius* strain DSM30822 and this clustering is strongly supported (aLRT value > 0.99), indicating that M9-20 is most closely related to *B. badius*. Also, isolates B9-9B and M9-1B are closely related to each other (aLRT value > 0.99) and are most closely related to *B. megaterium* strain ATCC145881 (aLRT value > 0.99).

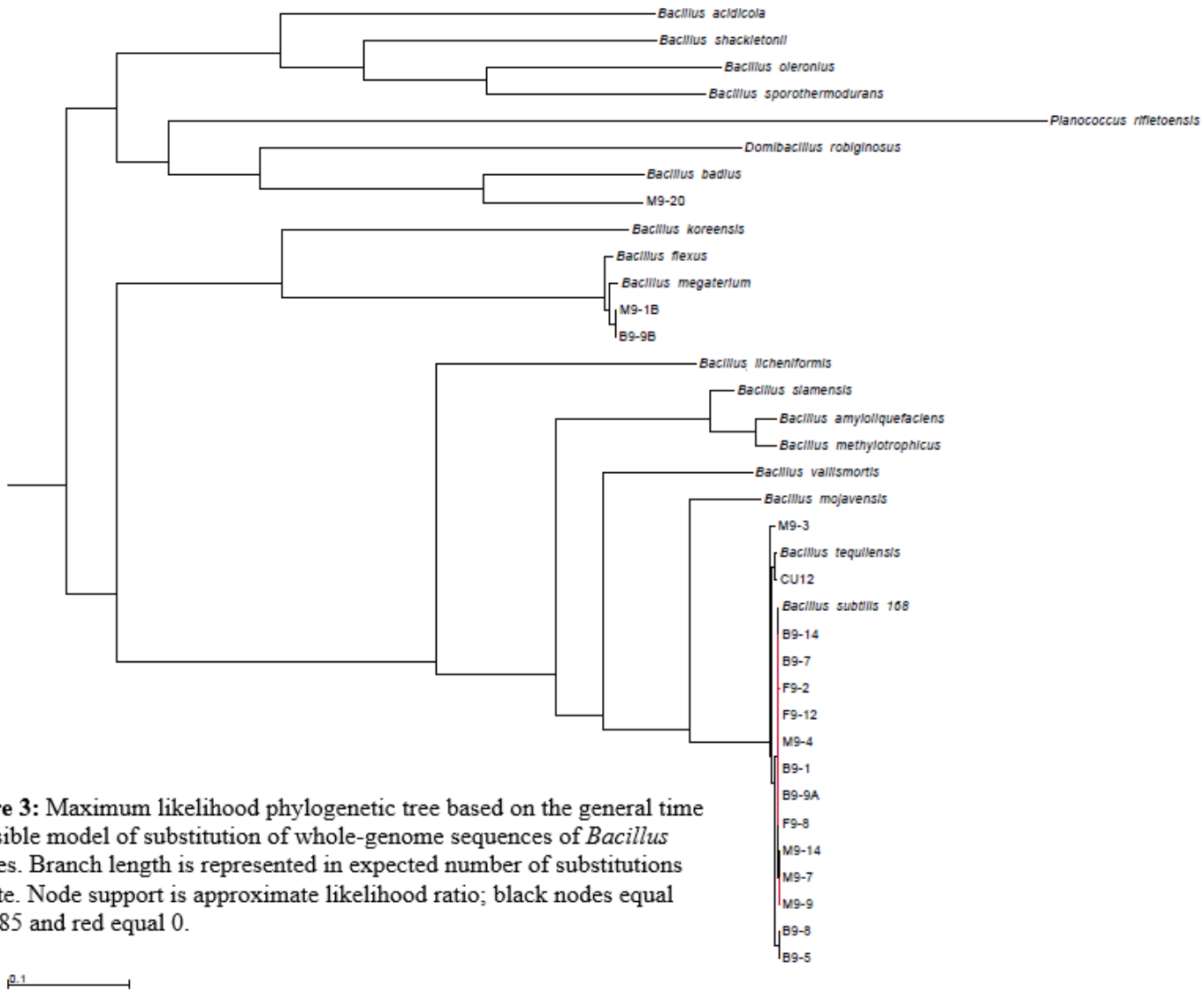

Figure 3: Maximum likelihood phylogenetic tree based on the general time reversible model of substitution of whole-genome sequences of *Bacillus* isolates. Branch length is represented in expected number of substitutions per site. Node support is approximate likelihood ratio; black nodes equal 0.99985 and red equal 0.

Looking at the section (subtree) of the *Bacillus* phylogenetic tree containing the *B. subtilis* isolates, isolate M9-3 branches off from the rest of the isolated samples, with the branching having strong support (aLRT value > 0.99). This isolate shares a common ancestor with *B. tequilensis* and *B. subtilis*, but it may represent a different species than the reference strains utilized to construct the phylogenetic tree. The clustering of isolate CU12 with *B. tequilensis* strain FJAT-14262a is highly supported (aLRT value > 0.99) and CU12 is most closely related with this reference strain. Isolates B9-5 and B9-8 branch off *B. subtilis* strain 168 (aLRT value > 0.99) and group together. They are highly related to each other, and are most closely related to *B. subtilis*. The remaining isolates (B9-1, B9-7, B9-9A, B9-14, F9-2, F9-8, F9-12, M9-4, M9-7, M9-9 and M9-14) all branch from *B. subtilis* strain 168 (aLRT value > 0.99) and all these isolates are most closely related to *B. subtilis*. There is branching within these isolates, separating M9-7, M9-9 and M9-14 from the rest (aLRT value > 0.99), and there is strong support for the branching of isolates M9-7 and M9-9 from M9-14 (aLRT value > 0.99). These three isolates are still highly related to the rest of the *B. subtilis* isolates.

When looking at the isolates branching from *B. subtilis* strain 168 (B9-1, B9-7, B9-9A, F9-2, F9-8, F9-12, and M9-4), the branch lengths are zero, indicating that there is essentially no evolutionary difference between these samples and they are all highly related to each other and to *B. subtilis*. These isolates also have node values of zero, indicating that there is no support for the structure of this part of the tree. There is no difference in likelihood scores around these nodes, which means the topology of this section can be rearranged and there would be no change in the relationships of the isolates or the likelihood scores.

3.3. BLAST analysis of cyclic lipopeptide biosynthesis genes

The cyclic lipopeptide biosynthesis genes were identified using the NCBI basic local alignment search tool (BLAST) on a local Linux machine, and these results were confirmed by performing a protein blast against reference proteins on the NCBI website. The genes searched for were genes for fengycin, plipastatin, surfactin and iturin.

Table 5 presents a summary of the nucleotide blast hits for the lipopeptide genes in each bacterial isolate used in this study. *Bacillus subtilis* isolates CU12, F9-8 and M9-9 revealed all the biosynthesis genes for all four CLPs (fengycin, plipastatin, surfactin and iturin). *Bacillus subtilis* isolates B9-1, B9-5, B9-7, B9-8, B9-9A, B9-14, F9-2, F9-12, M9-3, M9-4, M9-7 and M9-14 showed the biosynthesis genes for three CLPs (fengycin, plipastatin and surfactin). In the *B. subtilis* isolates, the genes for fengycin and plipastatin reside on the same stretch of DNA - *fenA* is very similar to *ppsD*, *fenB* is very similar to *ppsE*, *fenC* matches *ppsA*, *fenD* matches with *ppsB* and *fenE* matches *ppsC*. *Brevibacillus borstelensis* isolate M9-18 revealed the genes for fengycin and plipastatin, and 3 of the 4 genes for surfactin. *Bacillus badius* isolate M9-20 showed all five genes for fengycin, four of the five genes for plipastatin, but only two of four surfactin. *Bacillus megaterium* isolates B9-9B and M9-1B revealed two of the five genes for fengycin, one of the five for plipastatin and two of the four for surfactin.

Table 6 summarizes results for all nucleotide and protein blast results for the tested bacterial isolates. For *B. subtilis* isolates CU12 (Table A8), F9-8 (Table A10) and M9-9 (Table A16), nucleotide blasts indicated hits for sequences that matched the genes for fengycin (>79%), plipastatin (>98%), surfactin (>98%) and iturin (>70%), and the protein blasts produced results with high sequence homology (>95%) to nonribosomal peptide, plipastatin, surfactin and S-malonyltransferase and nonribosomal peptide synthases/synthetases, respectively.

For the remaining of the *B. subtilis* isolates (Tables A1-A5, A7, A9, A11, A13-15, and 17), nucleotide blasts indicated hits for sequences that matched the genes for fengycin (>79%), plipastatin (>96%) and surfactin (>98%), and the protein blasts produced results with high sequence homology (>80%) to nonribosomal peptide, plipastatin and surfactin synthases/synthetases, respectively (Table 6).

B. megaterium isolates B9-9B and M9-1B (Tables A6 and A12, respectively) showed results that were different than the *B. subtilis* isolates (Table 6). Indeed, nucleotide blasts indicated hits for sequences that matched two of the five genes for fengycin (>78%), one of the five genes for plipastatin (78%) and two of four for surfactin (>74%). The protein blasts produced results with high sequence homology (>97%) to acyl- (fengycin) and 2-succinylbenzoate CoA ligases (fengycin and plipastatin), as well as pyroglutamyl-peptidase I and long-chain fatty acid-CoA ligases for surfactin.

Nucleotide blasts for *B. badius* isolate M9-20 (Table A19) indicated hits for sequences that matched the genes for fengycin (>71%), four of the five genes for plipastatin (>80%) and three of four for surfactin (>79%), and the protein blasts produced results with high sequence homology (>79%) to o- and 2-succinylbenzoate CoA ligases for fengycin and plipastatin, and nonribosomal peptide synthetase, 2-succinylbenzoate-CoA ligase, and long-chain fatty acid-CoA ligase for surfactin (Table 6).

Finally, nucleotide blasts for *B. borstelensis* isolate M9-18 (Table A18) indicated hits for sequences that matched the genes for fengycin (>66%), plipastatin (>68%) and three of four for surfactin (>75%), and the protein blasts produced results with high sequence homology (>99%) to nonribosomal peptide synthetases for all three genes families (Table 6).

Table 5: Summary of nucleotide BLAST hits of CLP genes in bacterial isolates

Genes	fenA	fenB	fenC	fenD	fenE	srfAA	srfAB	srfAC	srfAD	ppsA	ppsB	ppsC	ppsD	ppsE	ituD	ituA	ituB	ituC
B9-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
B9-5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
B9-7	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
B9-8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
B9-9A	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
B9-9B	+	-	-	+	-	+	+	-	-	-	-	-	+	-	-	-	-	-
B9-14	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
CU12	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
F9-2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
F9-8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
F9-12	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
M9-1B	+	-	-	+	-	+	+	-	-	-	-	-	+	-	-	-	-	-
M9-3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
M9-4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
M9-7	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
M9-9	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
M9-14	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
M9-18	+	+	+	+	+	+	+	+	-	+	+	+	+	+	-	-	-	-
M9-20	+	+	+	+	+	+	+	-	-	+	-	+	+	+	-	-	-	-

Note: a plus sign indicates the presence of the gene and the dash indicates the absence of the gene; fenX are fengycin genes, srfXX are surfactin genes, ppsX are plipastatin genes and ituX are iturin genes.

Table 6: Summary of nucleotide and protein BLAST results for CLP genes in bacterial isolates[1]

	Fengycin		Surfactin		Plipastatin		Iturin	
	BLASTn	BLASTp	BLASTn	BLASTp	BLASTn	BLASTp	BLASTn	BLASTp
B9-1	>79%	NRPS[2]	>98%	surfactin synthetase	>96%	plipastatin synthetase	-	-
B9-5	>79%	NRPS	>98%	surfactin synthetase	>96%	plipastatin synthetase	-	-
B9-7	>79%	NRPS	>98%	surfactin synthetase	>96%	plipastatin synthetase	-	-
B9-8	>79%	NRPS	>98%	surfactin synthetase	>96%	plipastatin synthetase	-	-
B9-9A	>79%	NRPS	>98%	surfactin synthetase	>96%	plipastatin synthetase	-	-
B9-9B	>78%	acyl-CoA ligase	>74%	pyroglutamyl-peptidase and fatty acid-CoA ligases	78%	acyl- and 2-succinylbenzoate CoA ligase	-	-
B9-14	>79%	NRPS	>98%	surfactin synthetase	>96%	plipastatin synthetase	-	-
CU12	>79%	NRPS	>98%	surfactin synthetase	>98%	plipastatin synthetase	>70%	S-malonyltransferase and NRPS
F9-2	>79%	NRPS	>98%	surfactin synthetase	>96%	plipastatin synthetase	-	-
F9-8	>79%	NRPS	>98%	surfactin synthetase	>98%	plipastatin synthetase	>70%	S-malonyltransferase and NRPS
F9-12	>79%	NRPS	>98%	surfactin synthetase	>96%	plipastatin synthetase	-	-
M9-1B	>78%	acyl-CoA ligase	>74%	pyroglutamyl-peptidase and fatty acid-CoA ligases	78%	acyl- and 2-succinylbenzoate CoA ligase	-	-
M9-3	>79%	NRPS	>98%	surfactin synthetase	>96%	plipastatin synthetase	-	-
M9-4	>79%	NRPS	>98%	surfactin synthetase	>96%	plipastatin synthetase	-	-
M9-7	>79%	NRPS	>98%	surfactin synthetase	>96%	plipastatin synthetase	-	-
M9-9	>79%	NRPS	>98%	surfactin synthetase	>98%	plipastatin synthetase	>70%	S-malonyltransferase and NRPS
M9-14	>79%	NRPS	>98%	surfactin synthetase	>96%	plipastatin synthetase	-	-
M9-18	>66%	NRPS	>75%	NRPS	>68%	NRPS	-	-
M9-20	>71%	succinylbenzoate CoA ligases	>79%	succinylbenzoate- and fatty acid-CoA ligase and NRPS	>80%	succinylbenzoate CoA ligases	-	-

[1] Complete data available in appendixes (Tables A1-A19)

[2] NRPS = nonribosomal peptide synthetases

Chapter 4: Discussion

Chemical pesticides are widely used to control pests and reduce the amount of crop losses due to pest damage (Pimentel, 2009). However, these chemicals have many disadvantages, such as the possibility of pests developing resistance, the effects on human health from accidental poisonings and residues left on food, and the environmental damage, such as the reduced species diversity and soil contamination (Hart & Pimentel, 2002). Alternatives to these pesticides include biological control, which is the use of organisms to control pests, and this alternative is generally considered a safer method of controlling pests and reducing crop losses. Microorganisms that naturally antagonize pathogens can be found in many environments (Niisawa et al., 2008), and *Bacillus* spp. have been shown to be promising biocontrol agents that are effective at controlling fungal plant pathogens (Joseph, Ranjan Patra, & Lawrence, 2012). *Bacillus* bacteria produce many antimicrobial compounds that contribute to its ability to suppress plant diseases.

This study was carried out to confirm the identity of bacterial isolates antagonistic to plant pathogens that was determined using a phylogenetic analysis of the 16S gene of these isolates as well as to gain knowledge on the biosynthesis genes of cyclic lipopeptide produced by these bacterial antagonists.

4.1. Evaluation of phylogenetic relationships of *Bacillus* and *Brevibacillus* isolates

The identity of these bacterial isolates had been suggested in previous work using 16S rRNA sequences (Mohamed et al., 2017) through sequence matching with a curated ribosomal database and phylogenetic analyses by a maximum likelihood tree using the GTR model of substitution. In that work, sequence matching analyses determined that these isolates belonged to

the Phylum *Firmicutes*, in the *Bacilliales* order. The isolate M9-18 was shown to have very high similarity to *Brevibacillus borstelensis* strain DSM 6347^T (GenBank Accession No. AB112721) with 100% similarity value, and isolate M9-20 had been shown to be most similar (95.6%) to *Bacillus badius* strain ATCC 14574^T (X77790). The isolates B9-9B and M9-1B were shown to be most similar to *Bacillus megaterium* strain IAM 13418^T (D16273). Finally, fourteen isolates (B9-1, B9-5, B9-7, B9-8, B9-9A, B9-14, F9-2, F9-8, F9-12, M9-3, M9-4, M9-7, M9-9 and M9-14) were shown to be most similar (99.5-100%) to *Bacillus subtilis* strain DSM10^T (AJ276351). Phylogenetic analysis confirmed relatedness of the isolates to these taxonomic groups (Mohamed et al., 2017). Another study by Wise et al. (2012) indicated that a bacterium isolated from disease suppressive soil (named isolate CU12) has 16S rRNA sequences most similar to *B. subtilis*.

The phylogenetic analyses performed in this study using whole genome sequences provided results that were very similar to the results determined by 16S rRNA sequence analysis. Isolate M9-18 branched off the root of the tree with *Brevibacillus borstelensis* strain cifa_chp40 and away from the rest of the *Brevibacillus* reference genomes (Figure 2). This branching was highly supported (aLRT value > 0.99) and matched the branching in the 16S phylogeny found in Mohamed et al. (2017). Isolate M9-20 branched from *Domibacillus robinginosus* strain WS4628 and grouped with *Bacillus badius* strain DSM30822, which was highly supported (aLRT value > 0.99). This grouping also matched the branching in the 16S rRNA phylogeny. Also, isolates B9-9B and M9-1B grouped together (aLRT value > 0.99) and branched from *Bacillus megaterium* strain ATCC145881, resembling the structure of the 16S phylogenetic tree. Looking at the *Bacillus subtilis* subtree, isolate M9-3 branched off from the rest of the isolates, isolate CU12 grouped with *Bacillus tequilensis* strain FJAT-14262a and isolates B9-5 and B9-8 branched from *Bacillus subtilis* strain 168 and grouped together. Eleven isolates (B9-14, B9-7, F9-2, F9-12, M9-

4, B9-1, B9-9A, F9-8, M9-14, M9-7 and M9-9) branched from *Bacillus subtilis* strain 168 and were most similar to that reference strain. The branching for M9-3, CU12, B9-5 and B9-8 were highly supported (aLRT > 0.99).

These results generally matched expectations, as the identity of these isolates were determined by 16S phylogeny in previous work and so it was expected that the whole genome sequences would produce similar tree topologies. There are, however, some differences between the 16S and whole genome phylogenetic trees. First, isolate M9-3 was separated from the remainder of the *B. subtilis* isolates in the whole genome tree, whereas it was grouped with the *B. subtilis* isolates in the 16S tree. Also, the *B. subtilis* isolates were not all grouped together on one branch in the whole genome tree as they were in the 16S tree – they were separated and branch off each other. The node with the separation of M9-14, M9-7 and M9-9 from the other eight isolates was well supported, however the nodes representing the separation of B9-1, B9-9A, F9-8, B9-14, B9-7, F9-2, F9-12 and M9-4 possess an aLRT value of 0, which indicates that there were no differences in likelihood scores for topologies around the nodes. This means that the current topology of the tree around those nodes was the same as the other alternative topologies calculated by the approximate likelihood ratio test, and so the branching could be rearranged around those nodes and it would not impact the likelihood scores of the different topologies. The identity of isolate CU12 was previously determined to be *B. subtilis* using 16S gene sequences (Wise et al., 2012), however the whole genome phylogeny clustered it with *B. tequilensis*. Based on these results, it can be said that the identity of isolate M9-18 is *Brevibacillus borstelensis*, isolate M9-20 is *Bacillus badius*, isolates B9-9B and M9-1B are *Bacillus megaterium* and isolates B9-1, B9-5, B9-7, B9-8, B9-9A, B9-14, F9-2, F9-8, F9-12, M9-4, M9-7, M9-9 and M9-14 are *Bacillus subtilis*. The identity of isolate CU12 is *Bacillus tequilensis* and isolate M9-3 is closely related to

B. subtilis and *B. tequilensis*, however it may represent a different species than the reference strains utilized to construct the phylogenetic tree. More identification tests would have to be performed in order to assign M9-3 a species name.

The identification of a bacterial strain is an important part when studying prokaryotic taxonomy. Modern characterization of bacteria has been influenced by the advent of new genetic methods (Rappe & Giovannoni, 2003). The 16S rRNA gene is currently the basic tool for bacterial identification and characterization (Gomila et al., 2015). It is present in most bacteria and has a relatively slow evolution rate (Yamamoto et al., 2000). However, it is generally only effective at determining phylogenetic relationships to the genus level, and cannot distinguish between highly related species (Sentausa & Fournier, 2013). Another tool used for bacteria classification is multiple-locus sequence typing (MLST) and this method uses several housekeeping genes. Unfortunately, there is no standard set of genes that are used for this method and sequencing errors in the genomes can negatively impact the MLST results (Konstantinidis, Ramette, & Tiedje, 2006). The development of next-generation sequencing technologies has made sequence-based identification methods faster, more available and more cost-effective (Emerson, Agulto, Liu, & Liu, 2008). Phylogenetic analyses of bacteria using whole genome sequence data provides high resolution and removes the problems produced by sequence selection, compared to MLST and 16S rRNA analyses. Comparison using whole genome sequencing can allow for a more accurate identification of highly related bacterial species (Zhang et al., 2014). Whole genome sequencing data can provide valuable information that can be used for further tests and complement genomics-based methods, such as gene information and metabolite production (Emerson et al., 2008).

4.2. Evaluation of CLP biosynthesis genes

Bacillus and *Brevibacillus* isolates were tested to determine the presence of genes involved in the biosynthesis of *Bacillus* cyclic lipopeptides. Among the nineteen bacterial isolates tested, three isolates were found to possess sequences that matched the genes for the four tested CLPs, twelve isolates were found to possess sequences that matched the genes for three CLPs, one isolate were found to possess sequences that matched the genes for two CLPs, and three isolates were found to possess sequences that matched only some of the CLP biosynthesis genes.

When looking at the nucleotide blast results, sequences were found in the *B. subtilis* isolates that matched the fengycin genes (average percent identity is 80.2%), the surfactin genes (average percent identity is 99.2%), the plipastatin genes (average percent identity is 97.8%), and in three isolates the iturin genes (average percent identity is 75%). The sequences that matched the fengycin genes were located on the same positions of the genome as the sequences for the plipastatin genes – *fenA* is very similar to *ppsD*, *fenB* is very similar to *ppsE*, *fenC* matches *ppsA*, *fenD* matches with *ppsB* and *fenE* matches *ppsC*. Fengycin and plipastatin are nearly identical structurally – they differ only in the isomeric configuration of the amino acid tyrosine (fengycin has the amino acid in the D configuration, whereas plipastatin has it in the L configuration) (Eeman et al., 2009). Previous studies have shown that there is high homology between the two operons (Steller et al., 1999). Nucleotide sequences were found in the *B. megaterium* isolates that matched two of the fengycin genes (average percent identity is 81.5%), one of the plipastatin genes (percent identity is 78%) and two of the genes for surfactin (average percent identity is 79%). Nucleotide sequences were found in the *B. badius* isolate that matched the genes for fengycin (average percent identity is 80%), plipastatin (average percent identity is 81.7%) and three of the genes for surfactin (average percent identity is 82.3%). Nucleotide sequences were found in the *Brevibacillus*

borstelensis isolate that matched the genes for fengycin (average percent identity is 70.6%), plipastatin (average percent identity is 70.8%) and three of the genes for surfactin (average percent identity is 79.3%).

When looking the protein blast results, the nucleotide sequences in *B. subtilis* isolates that were found to match the fengycin, plipastatin, surfactin and iturin biosynthesis genes came back as nonribosomal peptide synthetases, plipastatin synthetase subunits, surfactin synthetase subunits and nonribosomal peptide synthetases, respectively. These results gave an average percent identity of 99.4%, 99.2%, 98.5% and 98.7%, respectively. The nucleotide sequences from *B. megaterium* isolates that matched the sequences for fengycin genes and one plipastatin gene came back as 2-succinylbenzoate-CoA ligase (*fenA* and *ppsD*) and acyl-COA ligase (*fenD*), with percent identity being 99% for each result. The sequences that matched genes from surfactin came back as pyroglutamyl-peptidase I (percent identity is 97%) and long-chain fatty acid-CoA ligase (percent identity is 99%). The nucleotide sequences from the *B. badius* isolate that matched sequences for fengycin genes *fenA*, *fenC* and *fenE* and plipastatin genes (*ppsA*, *ppsC* and *ppsD*) came back as 2-succinylbenzoate-CoA ligase (average percent identity is 83%) and sequences for fengycin genes *fenB* and *fenD* and plipastatin gene *ppsE* came back as o-succinylbenzoate-CoA ligase (average percent identity is 83.3%). The sequences that matched surfactin genes came back as nonribosomal peptide synthetase (percent identity 79%), long-chain fatty acid-CoA ligase (percent identity 84%) and 2-succinylbenzoate-CoA ligase (percent identity 82%). The nucleotide sequences that were found to match the fengycin, plipastatin, and surfactin biosynthesis genes came back as nonribosomal peptide synthetases (average percent identity is 99% for all genes).

For the *B. subtilis* isolates and the *Brevibacillus borstelensis* isolate, the nucleotide sequences that matched the cyclic lipopeptide genes provided protein blast results of nonribosomal

peptide synthetases/synthases, which are large multienzyme complexes that incorporate amino acids into a peptide (Marahiel, Stachelhaus, & Mootz, 1997; Stein et al., 1996; Walsh, 2008). The sequences from the *B. megaterium* isolates came back with 2-succinylbenzoate-CoA ligases, which are enzymes involved in the synthesis of menaquinone, acyl-CoA ligases, which are involved in acetate metabolism, pyroglutamyl-peptidase I, which is an enzyme that catalyzes the release of a pyroglutamyl group form a polypeptide, and long-chain fatty acid-CoA ligases, which are enzymes involved in the breakdown of long-chain fatty acids. The sequences for the *B. badius* isolate got protein blast results of 2- and o-succinylbenzoate-CoA ligases, long-chain fatty acid-CoA ligase and a nonribosomal peptide synthetase. These results, together with the percent identity of both the nucleotide and protein blasts, indicate that the CLP operons for fengycin/plipastatin and surfactin were present in the *B. subtilis* isolates and the *Brevibacillus borstelensis* isolate, and the iturin operon was present in CU12, F9-8 and M9-9. These results also indicate that the CLP operons for fengycin/plipastatin, surfactin and iturin were not present in the *B. badius* and *B. megaterium* isolates.

There are discrepancies between the nucleotide identities and the protein identities for the fengycin genes of the *B. subtilis* isolates. Nucleotide sequences for sequences matching fengycin genes give an average percent identity of 80.2%, but the protein sequences gave an average percent identity of 99.4%. This difference in percent identity could potentially come from the *B. subtilis* isolates taking in the fengycin genes from other strains or species, as *Bacillus* is known to be good at taking up exogenous DNA (Dubnau, 1991). *B. subtilis* has developed many strategies to survive in a wide variety of environments, and when the bacterium is exposed to environmental stress, it transforms into competent cells, which then are able to take in DNA from the environment (Hamoen, Venema, & Kuipers, 2003) Also, protein coding sequences are more conserved than

nucleotide sequences, and so if there were synonymous mutations that occurred in the nucleotide sequence, the protein sequence would remain the same.

Previous work (Mohamed et al., 2017) determined the lipopeptide families produced by these antagonistic bacterial isolates. The lipopeptides were extracted from liquid culture and analyzed using liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS). The mass spectra produced were compared to authentic standards and MS data presented in the literature to identify the lipopeptides (results summarized in Table 7).

As shown in Table 7, seven *B. subtilis* isolates (B9-1, B9-5, B9-7, B9-14, M9-4, M9-7, M9-14) and *Brevibacillus borstelensis* isolate M9-18 possess the genes for fengycin/plipastatin and surfactin, and these isolates produce those cyclic lipopeptides. Also, *B. subtilis* isolates CU12 and M9-9 possess the genes for fengycin/plipastatin, surfactin and iturin and these isolates have been shown to produce these compounds. In addition, five *B. subtilis* isolates (B9-8, B9-9A, F9-2, F9-12, M9-3) were shown to produce only surfactin, and *B. subtilis* F9-8 was shown to produce surfactin and iturin. These isolates did possess the corresponding CLP genes, but also possess the genes for fengycin, which none had been shown to produce in previous work (Mohamed et al., 2017).

Table 7: Cyclic lipopeptides produced and presence of CLP genes by bacterial isolates

Isolate	Fengycin/plipastatin		Iturin		Surfactin	
	Product[1]	Genes present[2]	Product[1]	Genes present[2]	Product[1]	Genes present[2]
B. subtilis B9-1	yes	yes	no	no	yes	yes
B. subtilis B9-5	yes	yes	no	no	yes	yes
B. subtilis B9-7	yes	yes	no	no	yes	yes
B. subtilis B9-8	no	yes	no	no	yes	yes
B. subtilis B9-9A	no	yes	no	no	yes	yes
B. subtilis B9-14	yes	yes	no	no	yes	yes
B. subtilis F9-2	no	yes	no	no	yes	yes
B. subtilis F9-8	yes	yes	yes	yes	yes	yes
B. subtilis F9-12	no	yes	no	no	yes	yes
B. subtilis M9-3	no	yes	no	no	yes	yes
B. subtilis M9-4	yes	yes	no	no	yes	yes
B. subtilis M9-7	yes	yes	no	no	yes	yes
B. subtilis M9-9	yes	yes	yes	yes	yes	yes
B. subtilis M9-14	yes	yes	no	no	yes	yes
B. megaterium B9-9B	yes	no	no	no	yes	no
B. megaterium M9-1B	yes	no	no	no	yes	no
B. badius M9-20	yes	no	no	no	yes	no
Brevibacillus borstelensis M9-18	yes	yes	no	no	yes	yes

[1] Adapted from Mohamed et al. (2017)

[2] This study

Shading represents different gene/product combination; white indicates the genes and products are present, light grey indicates that genes present, but at least one product is not present, and dark grey indicates no genes, but product present.

A possible reason for why fengycin was produced in these isolates may be that there is a problem with the *fenp* promoter. It was determined that there is a promoter (*fenp*) located upstream from the first fengycin gene (*fenC*) and this promoter is responsible for the transcription of the fengycin operon (Lin, Chen, Chang, Tschen, & Liu, 1999). For transcription to occur, RNA polymerase needs to bind to the promoter and *fenp* contains an UP element, which assists with RNA polymerase binding to the promoter. Studies have shown that *fenp* is required for the fengycin synthetase operon to be transcribed and for fengycin to be produced (Ke, Chang, Lin, & Liu, 2009). There could also be an additional regulatory reason for why fengycin is not being produced. There could be a stop codon present in the sequence, which could have occurred because of a mutation. A nonsense mutation in the nucleotide sequence could have changed an amino acid to a stop codon, which would have prevented transcription of the biosynthesis genes and the production of fengycin. The open reading frame of the gene sequences could be analyzed to look for these types of mutations.

Another reason why these isolates may not have produced fengycin is that there are important culture parameters that need to be taken into account for the production of lipopeptides by *B. subtilis*. Parameters such as growing temperature, pH of the medium, concentration of peptone, phosphate and yeast extract in the medium and shaking, all play a major role in the lipopeptide production. A Plackett-Burman experimental design was used to identify the important factors in lipopeptide production and determine the optimal conditions (Jacques et al., 1999). This optimization allowed the authors to find conditions adapted for lipopeptide production: temperature, 30°C; pH, 7.0; shaking, 200 rpm; sucrose, 20 g/L; yeast extract: 7 g/L peptone, 30 g/L; KH_2PO_4, 1.9 g/L; $MnSO_4$ and $MgSO_4$, 9 mL/L; trace elements, 1 mL/L. These results correspond to *B. subtilis* being grown in a rich culture medium, and this was confirmed by

another study that looked at the effects of different amino acids on the production of lipopeptide homologs (Akpa et al., 2001). Another interesting result regarding the effect of shaking on lipopeptide production is that it was shown in a few studies the effect oxygen has on the biosynthesis of lipopeptides. It was shown that the production of fengycin was favorable when larger flasks were used and there were lower shaking frequencies (Fahim et al., 2012). Also, fengycin synthesis was higher when there were moderate oxygenation conditions, which can be obtained with various shaking frequencies and flask sizes (Guez, Muller, Danze, Buchs, & Jacques, 2008; Hbid et al., 1996).

It is also possible that the fengycin genes within these isolates were truncated or had deleted sequences. A traditional nucleotide BLAST was performed in order to identify the cyclic lipopeptide genes, and the program may have identified the genes as being present, even if a relatively large amount of the sequence was absent. Work could be done to perform a protein truncated test (PTT), which could identify if there was a mutation in the sequence that results in truncation. The specific region of DNA encoding the protein would be isolated and amplified, transcription and translation of the protein would be performed in vivo and the translated protein would be identified on an electrophoresis gel. The mutated/truncated product would be easily distinguished from the full-sized protein (Den Dunnen & Van Ommen, 1999). In addition to this, other programs capable of predicting bacterial genes could be used, such as Glimmer (Delcher, Bratke, Powers, & Salzberg, 2007), GenemarkHMM (Lukashin & Borodovsky, 1998) and Prodigal (Hyatt et al., 2010).

Previous work had determined the lipopeptides produced by each bacterial isolate (Mohamed et al., 2017) extracted lipopeptides from *Bacillus* cultures that were grown in 50 mL of medium optimized for lipopeptide production (MOLP) (Akpa et al., 2001) at 30°C shaking (120

rev/min) for 72 hours. These conditions may not have been ideal for the production of fengycin in five of the *B. subtilis* isolates, which may have been the cause of fengycin not being produced. In addition, these lipopeptides are produced when the organism is under stress (Reder-Christ et al., 2012), and the medium may not have been sufficient to induce fengycin production in those isolates. In previous work, *B. subtilis* isolate B9-5 was grown in the optimized medium with fungal pathogens (*Alternaria solani*, *Fusarium sambucinum*, *Verticillium dahliae*, and *Rhizopus stolonifer*) and it was shown that there was an increase in total fengycin homologs when *B. subtilis* B9-5 was grown in the presence of a competitive mold, *V. dahliae* (Groulx, 2015).

Previous work was also done to analyze and quantify the production of lipopeptides produced by *B. subtilis* over time. One study identified the lipopeptides produced by *B. subtilis* strains during each growth phase, and found that lipopeptides were produced at the start and halfway through the log growth phase (Dunlap, Schisler, Price, & Vaughn, 2011). Another study performed time-course studies to quantify the amount of lipopeptides produced by *B. subtilis* strain B9-5. It was determined that the maximum concentration of fengycin was produced at 48 hours, following the stationary growth phase. The *B. subtilis* isolates B9-8, B9-9A, F9-2, F9-12 and M9-3 were grown for 72 hours, so there may have been a decrease or degradation of fengycin after 48 hours, which could contribute to fengycin not being found in those isolates (DeFilippi, 2016) even when the genes were present.

It was shown that *B. megaterium* isolates B9-9B and M9-1B, along with *B. badius* isolate M9-20 produce fengycin and surfactin, however it was determined that these isolates did not possess those gene operons. The protein blasts of the sequences matching the lipopeptide genes produced results of enzymes that were not involved in cyclic lipopeptide biosynthesis. A possible reason for this could be that the cyclic lipopeptide genes have not been properly identified in these

species yet. There have been previous studies showing the production of surfactin (Rangarajan, Dhanarajan, Kumar, Sen, & Mandal, 2012) and fengycin (Rangarajan, Dhanarajan, & Sen, 2015) in *B. megaterium*, however these studies only identified the compounds using high performance thin layer chromatography and Fourier transform infrared spectrometry, no genetic tests or PCR was performed to identify biosynthesis genes. The nucleotide sequences of lipopeptide genes used in the blasts were from *B. subtilis* and it may not be possible to produce a blast hit on a nucleotide sequence that has not been identified in the species being tested. It was not possible to find lipopeptide gene sequences belonging to *B. megaterium* or *B. badius* on NCBI. No studies could be found on the production of cyclic lipopeptides by *B. badius*. It may also be necessary to reanalyze bacterial extracts using liquid chromatography-mass spectrometry to confirm production of lipopeptides in these isolates.

Another reason for the *B. badius* and *B. megaterium* isolates not possessing the genes for fengycin/plipastatin and surfactin could be that there were issues with the assemblies. The coverage of the assemblies varies between 34 and 55 times per base. It is possible that the coverage was not high enough and the sequences of the cyclic lipopeptide genes did not make it into the genome assembly. A possible solution for this would be to construct genome assemblies of these bacterial isolates using a reference-based assembler and to use a reference strain of *B. badius* and *B. megaterium* that has been shown in the literature to produce cyclic lipopeptides. This way, there is a higher chance of the biosynthesis genes getting incorporated into the assembly and being identified. In addition to this, the sequence reads that did not get incorporated into the genome assemblies may potentially contain the cyclic lipopeptide gene sequences. A nucleotide BLAST could be performed with the leftover sequence reads to determine if the genes are present in those sequences. It is also possible that there was contamination in the *B. badius* and *B. megaterium*

isolates. These cultures could have been contaminated with a *B. subtilis* culture, and when the *B. subtilis* culture started producing cyclic lipopeptides, it was mistaken for *B. badius* and *B. megaterium*. It would be a good idea to start a new culture of these bacterial isolates and test for the production of these cyclic lipopeptides again, to ensure that these isolates in fact produce cyclic lipopeptides.

4.3. Future work

Follow-up work on this project could include analysis of the expression of fengycin genes in five *B. subtilis* isolates (B9-8, B9-9A, F9-2, F9-12 and M9-3) using real-time reverse transcription polymerase chain reaction (RT-PCR), which measures the amount of RNA expression. Also, time-course studies could be performed with those isolates to quantify the amount of lipopeptides produced over time and determine if fengycin is produced at a sooner or later time point. Plackett-Burman experimental designs could be performed to determine the optimal conditions for fengycin production in these isolates. These isolates could be cultured in liquid medium with various fungal pathogens to determine if the presence of a competitor would be sufficient to stress the bacteria and induce production of fengycin.

With genomic sequences now available, the antagonistic effect of these bacteria could be further studied by searching for genes responsible for the production of other antimicrobial compounds. *Bacillus* species are known to produce secondary metabolites other than peptide antibiotics, such as polyketides, plant growth hormones and terpenoids (Hamdache, Lamarti, Aleu, & Collado, 2011). The genomic information also allows for the study of mobile genetic elements such as transposons. These DNA sequences move around in the genome and they often carry

antibiotic resistance genes (Burrus & Waldor, 2004). It would be interesting to analyze the genomes of these bacterial isolates to see if they possess any mobile elements and antibiotic resistance genes as it has been documented that *B. subtilis* contain these elements (Auchtung, Lee, Monson, Lehman, & Grossman, 2005).

With the information from the phylogenetic analyses and the genome sequences, the distribution of genes within these bacteria can be studied such as the acquisition and loss of genes. Bacterial genomes are constantly in flux when it comes to size and gene content (X. Didelot, Darling, & Falush, 2009). Work could be done to analyze the protein coding sections of these bacterial genomes to look at the gain and loss of genes over time, what types of genes stay conserved over time and if there is a relationship between gain, loss, and environment.

When the phylogenetic analyses were performed, recombination was not something that was taken into consideration. Bacteria undergo mechanisms of recombination, which replaces segments of DNA with that of another organism's and if this is not accounted for in the phylogeny, it can provide misleading representation of the relationships between bacterial species (Xavier Didelot & Wilson, 2015). Further work on running a phylogenetic analysis that incorporates recombination events could be done, and then this analysis could be compared to the original phylogenetic analysis to determine discrepancies.

Chapter 5: Conclusion

The whole genome phylogeny confirmed the identity of thirteen *B. subtilis* isolates, two *B. megaterium* isolates, one *B. badius* isolate and one *Brevibacillus borstelensis* isolate. Isolate CU12 was identified as *B. tequilensis*, which was different than the identity assigned using 16S rRNA

sequences, and isolate M9-3 was clustered on its own, indicating that it may be a different species. The nucleotide and protein blasts identified the presence of cyclic lipopeptide gene operons in the *B. subtilis* and *B. borstelensis* isolates, matching production data determined in previous work, and operon were not identified in the *B. megaterium* and *B. badius* isolates. This study gives insight into the genetic mechanisms behind cyclic lipopeptide production and future work is required to study the expression of these genes.

Chapter 6: References

Akpa, E., Jacques, P., Wathelet, B., Paquot, M., Fuchs, R., Budzikiewicz, H., & Thonart, P. (2001). Influence of culture conditions on lipopeptide production by *Bacillus subtilis*. *Applied Biochemistry and Biotechnology, 91-3*, 551-561. doi:10.1385/abab:91-93:1-9:551

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403-410. doi:http://dx.doi.org/10.1016/S0022-2836(05)80360-2

Arrebola, E., Jacobs, R., & Korsten, L. (2010). Iturin A is the principal inhibitor in the biocontrol activity of *Bacillus amyloliquefaciens* PPCB004 against postharvest fungal pathogens. *Journal of Applied Microbiology, 108*(2), 386-395. doi:10.1111/j.1365-2672.2009.04438.x

Auchtung, J. M., Lee, C. A., Monson, R. E., Lehman, A. P., & Grossman, A. D. (2005). Regulation of a Bacillus subtilis mobile genetic element by intercellular signaling and the global DNA damage response. *Proceedings of the National Academy of Sciences of the United States of America, 102*(35), 12554-12559. doi:10.1073/pnas.0505835102

Bais, H. P., Fall, R., & Vivanco, J. M. (2004). Biocontrol of *Bacillus subtilis* against infection of Arabidopsis roots by *Pseudomonas syringae* is facilitated by biofilm formation and surfactin production. *Plant Physiology, 134*(1), 307-319. doi:10.1104/pp.103.028712

Bernier-English, V., Avis, T. J., Mimee, B., Antoun, H., & Tweddell, R. J. (2010). Compost amendment, a potential alternative to soil fumigation for the control of strawberry verticillium wilt. *Phytopathology, 100*(6), S193-S193.

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., . . . Xia, F. (2015). RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports, 5*, 8365. doi:10.1038/srep08365

Burrus, V., & Waldor, M. K. (2004). Shaping bacterial genomes with integrative and conjugative elements. *Research in Microbiology, 155*(5), 376-386. doi:10.1016/j.resmic.2004.01.012

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics, 10*, 421-421. doi:10.1186/1471-2105-10-421

Cane, D. E., Walsh, C. T., & Khosla, C. (1998). Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science, 282*(5386), 63.

Carrillo, C., Teruel, J. A., Aranda, F. J., & Ortiz, A. (2003). Molecular mechanism of membrane permeabilization by the peptide antibiotic surfactin. *Biochimica et Biophysica Acta (BBA) - Biomembranes, 1611*(1–2), 91-97. doi:https://doi.org/10.1016/S0005-2736(03)00029-4

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution, 17*(4), 540-552. doi:10.1093/oxfordjournals.molbev.a026334

Cosmina, P., Rodriguez, F., Deferra, F., Grandi, G., Perego, M., Venema, G., & Vansinderen, D. (1993). Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. *Molecular Microbiology, 8*(5), 821-831. doi:10.1111/j.1365-2958.1993.tb01629.x

Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Research, 14*(7), 1394-1403. doi:10.1101/gr.2289704

Delcher, A. L., Bratke, K. A., Powers, E. C., & Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics, 23*(6), 673-679. doi:10.1093/bioinformatics/btm009

Deleu, M., Paquot, M., & Nylander, T. (2008). Effect of fengycin, a lipopeptide produced by *Bacillus subtilis*, on model biomembranes. *Biophysical Journal, 94*(7), 2667-2679. doi:10.1529/biophysj.107.114090

Den Dunnen, J. T., & Van Ommen, G. J. B. (1999). The protein truncation test: A review. *Human Mutation, 14*(2), 95-102. doi:10.1002/(sici)1098-1004(1999)14:2<95::aid-humu1>3.0.co;2-g

Didelot, X., Darling, A., & Falush, D. (2009). Inferring genomic flux in bacteria. *Genome Research, 19*(2), 306-317. doi:10.1101/gr.082263.108

Didelot, X., & Wilson, D. J. (2015). ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLOS Computational Biology, 11*(2), e1004041. doi:10.1371/journal.pcbi.1004041

Dubnau, D. (1991). Genetic competence in *Bacillus subtilis*. *Microbiological Reviews, 55*(3), 395-424.

Dunlap, C. A., Schisler, D. A., Price, N. P., & Vaughn, S. F. (2011). Cyclic lipopeptide profile of three *Bacillus subtilis* strains; antagonists of Fusarium head blight. *Journal of Microbiology, 49*(4), 603-609. doi:10.1007/s12275-011-1044-y

Eeman, M., Pegado, L., Dufrêne, Y. F., Paquot, M., & Deleu, M. (2009). Influence of environmental conditions on the interfacial organisation of fengycin, a bioactive lipopeptide produced by Bacillus subtilis. *Journal of Colloid and Interface Science, 329*(2), 253-264. doi:http://dx.doi.org/10.1016/j.jcis.2008.10.017

Emerson, D., Agulto, L., Liu, H., & Liu, L. P. (2008). Identifying and characterizing bacteria in an era of genomics and proteomics. *Bioscience, 58*(10), 925-936. doi:10.1641/b581006

Fahim, S., Dimitrov, K., Gancel, F., Vauchel, P., Jacques, P., & Nikov, I. (2012). Impact of energy supply and oxygen transfer on selective lipopeptide production by Bacillus subtilis BBG21. *Bioresource Technology, 126*, 1-6. doi:10.1016/j.biortech.2012.09.019

Glare, T., Caradus, J., Gelernter, W., Jackson, T., Keyhani, N., Kohl, J., . . . Stewart, A. (2012). Have biopesticides come of age? *Trends in Biotechnology, 30*(5), 250-258. doi:10.1016/j.tibtech.2012.01.003

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., . . . Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. *Science, 327*(5967), 812.

Gomila, M., Peña, A., Mulet, M., Lalucat, J., & García-Valdés, E. (2015). Phylogenomics and systematics in *Pseudomonas*. *Frontiers in Microbiology, 6*, 214. doi:10.3389/fmicb.2015.00214

Gonzalez, D. J., Haste, N. M., Hollands, A., Fleming, T. C., Hamby, M., Pogliano, K., . . . Dorrestein, P. C. (2011). Microbial competition between *Bacillus subtilis* and

Staphylococcus aureus monitored by imaging mass spectrometry. *Microbiology, 157*(Pt 9), 2485-2492. doi:10.1099/mic.0.048736-0

Grover, M., Nain, L., Singh, S. B., & Saxena, A. K. (2010). Molecular and biochemical approaches for characterization of antifungal trait of a potent biocontrol agent *Bacillus subtilis* RP24. *Current Microbiology, 60*(2), 99-106. doi:10.1007/s00284-009-9508-6

Guez, J. S., Muller, C. H., Danze, P. M., Buchs, J., & Jacques, P. (2008). Respiration activity monitoring system (RAMOS), an efficient tool to study the influence of the oxygen transfer rate on the synthesis of lipopeptide by *Bacillus subtilis* ATCC6633. *Journal of Biotechnology, 134*(1-2), 121-126. doi:10.1016/j.jbiotec.2008.01.003

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology, 59*(3), 307-321. doi:10.1093/sysbio/syq010

Hamdache, A., Lamarti, A., Aleu, J., & Collado, I. G. (2011). Non-peptide metabolites from the genus *Bacillus*. *Journal of Natural Products, 74*(4), 893-899. doi:10.1021/np100853e

Hamoen, L. W., Venema, G., & Kuipers, O. P. (2003). Controlling competence in *Bacillus subtilis*: shared use of regulators. *Microbiology-Sgm, 149*, 9-17. doi:10.1099/mic.0.26003-0

Hart, K., & Pimentel, D. (2002). Public health and costs of pesticides. *Encyclopedia of Pest Management, New York, Marcel Dekker*, 677-679.

Harwood, C. R. (1992). *Bacillus subtilis* and its relatives - molecular biological and industrial workhorses. *Trends in Biotechnology, 10*(7), 247-256. doi:10.1016/0167-7799(92)90233-l

Hbid, C., Jacques, P., Razafindralambo, H., Mpoyo, M. K., Meurice, E., Paquot, M., & Thonart, P. (1996). Influence of the production of two lipopeptides, Iturin A and Surfactin S1, on oxygen transfer during *Bacillus subtilis* fermentation. *Applied Biochemistry and Biotechnology, 57-8*, 571-579. doi:10.1007/bf02941737

Heerklotz, H., Wieprecht, T., & Seelig, J. (2004). Membrane perturbation by the lipopeptide surfactin and detergents as studied by Deuterium NMR. *The Journal of Physical Chemistry B, 108*(15), 4909-4915. doi:10.1021/jp0371938

Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *Bmc Bioinformatics, 11*, 11. doi:10.1186/1471-2105-11-119

Jacques, P., Hbid, C., Destain, J., Razafindralambo, H., Paquot, M., De Pauw, E., & Thonart, P. (1999). Optimization of biosurfactant lipopeptide production from *Bacillus subtilis* S499 by Plackett-Burman design. *Applied Biochemistry and Biotechnology, 77-9*, 223-233. doi:10.1385/abab:77:1-3:223

Janda, J. M., & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of Clinical Microbiology, 45*(9), 2761-2764. doi:10.1128/JCM.01228-07

Joseph, B., Ranjan Patra, R., & Lawrence, R. (2012). Characterization of plant growth promoting rhizobacteria associated with chickpea (*Cicer arietinum L.*). *International Journal of Plant Production, 1*(2), 141-152. doi:10.22069/ijpp.2012.532

Kavroulakis, N., Ntougias, S., Besi, M. I., Katsou, P., Damaskinou, A., Ehaliotis, C., . . . Papadopoulou, K. K. (2010). Antagonistic bacteria of composted agro-industrial residues exhibit antibiosis against soil-borne fungal plant pathogens and protection of tomato plants

from *Fusarium oxysporum f.sp radicis-lycopersici*. *Plant and Soil, 333*(1-2), 233-247. doi:10.1007/s11104-010-0338-x

Ke, W. J., Chang, B. Y., Lin, T. P., & Liu, S. T. (2009). Activation of the promoter of the fengycin synthetase operon by the UP element. *Journal of Bacteriology, 191*(14), 4615-4623. doi:10.1128/jb.00255-09

Konstantinidis, K. T., Ramette, A., & Tiedje, J. M. (2006). Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. *Applied and Environmental Microbiology, 72*(11), 7286-7293. doi:10.1128/aem.01398-06

Kopp, F., & Marahiel, M. A. (2007). Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. *Natural Product Reports, 24*(4), 735-749. doi:10.1039/B613652B

Lin, T. P., Chen, C. L., Chang, L. K., Tschen, J. S. M., & Liu, S. T. (1999). Functional and transcriptional analyses of a fengycin synthetase gene, fenC, from *Bacillus subtilis*. *Journal of Bacteriology, 181*(16), 5060-5067.

Liu, J. J., Hagberg, I., Novitsky, L., Hadj-Moussa, H., & Avis, T. J. (2014). Interaction of antimicrobial cyclic lipopeptides from *Bacillus subtilis* influences their effect on spore germination and membrane permeability in fungal plant pathogens. *Fungal Biology, 118*(11), 855-861. doi:10.1016/j.funbio.2014.07.004

Lukashin, A. V., & Borodovsky, M. (1998). GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Research, 26*(4), 1107-1115. doi:10.1093/nar/26.4.1107

Marahiel, M. A., Stachelhaus, T., & Mootz, H. D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chemical Reviews, 97*(7), 2651-2673. doi:10.1021/cr960029e

Martin-Lapierre, A., Mimee, B., & Tweddell, R. J. (2011). Evaluation of combined effect of compost amendments and fumigation on strawberry verticillium wilt. *Phytopathology, 101*(6), S261-S261.

Mohamed, R., Groulx, E., Defilippi, S., Erak, T., Tambong, J. T., Tweddell, R. J., . . . Avis, T. J. (2017). Physiological and molecular characterization of compost bacteria antagonistic to soil-borne plant pathogens. *Canadian Journal of Microbiology, 63*(5), 411-426. doi:10.1139/cjm-2016-0599

Moszer, I., Jones, L. M., Moreira, S., Fabry, C., & Danchin, A. (2002). SubtiList: the reference database for the *Bacillus subtilis* genome. *Nucleic Acids Research, 30*(1), 62-65. doi:10.1093/nar/30.1.62

Niisawa, C., Oka, S.-i., Kodama, H., Hirai, M., Kumagai, Y., Mori, K., . . . Miyamoto, H. (2008). Microbial analysis of a composted product of marine animal resources and isolation of bacteria antagonistic to a plant pathogen from the compost. *The Journal of General and Applied Microbiology, 54*(3), 149-158. doi:10.2323/jgam.54.149

Oerke, E. C. (2005). Crop losses to pests. *The Journal of Agricultural Science, 144*(1), 31-43. doi:10.1017/S0021859605005708

On, A. N., Wong, F., Ko, Q., Tweddell, R. J., Antoun, H., & Avis, T. J. (2015). Antifungal effects of compost tea microorganisms on tomato pathogens. *Biological Control, 80*, 63-69. doi:10.1016/j.biocontrol.2014.09.017

Ongena, M., & Jacques, P. (2008). *Bacillus* lipopeptides: versatile weapons for plant disease biocontrol. *Trends in Microbiology, 16*(3), 115-125. doi:https://doi.org/10.1016/j.tim.2007.12.009

Ongena, M., Jacques, P., Touré, Y., Destain, J., Jabrane, A., & Thonart, P. (2005). Involvement of fengycin-type lipopeptides in the multifaceted biocontrol potential of *Bacillus subtilis*. *Applied Microbiology and Biotechnology, 69*(1), 29. doi:10.1007/s00253-005-1940-3

Patel, H., Tscheka, C., Edwards, K., Karlsson, G., & Heerklotz, H. (2011). All-or-none membrane permeabilization by fengycin-type lipopeptides from Bacillus subtilis QST713. *Biochimica et Biophysica Acta (BBA) - Biomembranes, 1808*(8), 2000-2008. doi:https://doi.org/10.1016/j.bbamem.2011.04.008

Patel, J. B. (2001). 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *Molecular Diagnosis, 6*(4), 313-321. doi:10.1054/modi.2001.29158

Perez-Garcia, A., Romero, D., & de Vicente, A. (2011). Plant protection and growth stimulation by microorganisms: biotechnological applications of Bacilli in agriculture. *Current Opinion in Biotechnology, 22*(2), 187-193. doi:10.1016/j.copbio.2010.12.003

Peypoux, F., Bonmatin, J. M., & Wallach, J. (1999). Recent trends in the biochemistry of surfactin. *Applied Microbiology and Biotechnology, 51*(5), 553-563. doi:10.1007/s002530051432

Pimentel, D. (2009). Pesticides and pest control. In *Integrated pest management: innovation-development process* (pp. 83-87): Springer.

Raaijmakers, J. M., De Bruijn, I., Nybroe, O., & Ongena, M. (2010). Natural functions of lipopeptides from *Bacillus* and *Pseudomonas*: more than surfactants and antibiotics. *FEMS Microbiology Reviews, 34*(6), 1037-1062. doi:10.1111/j.1574-6976.2010.00221.x

Ramarathnam, R., Bo, S., Chen, Y., Fernando, W. G. D., Xuewen, G., & de Kievit, T. (2007). Molecular and biochemical detection of fengycin- and bacillomycin D-producing *Bacillus* spp., antagonistic to fungal pathogens of canola and wheat. *Canadian Journal of Microbiology, 53*(7), 901-911. doi:10.1139/W07-049

Rangarajan, V., Dhanarajan, G., Kumar, R., Sen, R., & Mandal, M. (2012). Time-dependent dosing of Fe2+ for improved lipopeptide production by marine *Bacillus megaterium*. *Journal of Chemical Technology and Biotechnology, 87*(12), 1661-1669. doi:10.1002/jctb.3814

Rangarajan, V., Dhanarajan, G., & Sen, R. (2015). Bioprocess design for selective enhancement of fengycin production by a marine isolate *Bacillus megaterium*. *Biochemical Engineering Journal, 99*, 147-155. doi:10.1016/j.bej.2015.03.016

Rappe, M. S., & Giovannoni, S. J. (2003). The uncultured microbial majority. *Annual Review of Microbiology, 57*, 369-394. doi:10.1146/annurev.micro.57.030502.090759

Reder-Christ, K., Schmidt, Y., Dorr, M., Sahl, H. G., Josten, M., Raaijmakers, J. M., . . . Bendas, G. (2012). Model membrane studies for characterization of different antibiotic activities of lipopeptides from *Pseudomonas*. *Biochimica Et Biophysica Acta-Biomembranes, 1818*(3), 566-573. doi:10.1016/j.bbamem.2011.08.007

Robinson, D. F., & Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences, 53*(1), 131-147. doi:http://dx.doi.org/10.1016/0025-5564(81)90043-2

Romero, D., de Vicente, A., Rakotoaly, R. H., Dufour, S. E., Veening, J.-W., Arrebola, E., . . . Pérez-García, A. (2007). The iturin and fengycin families of lipopeptides are key factors in antagonism of *Bacillus subtilis* toward *Podosphaera fusca*. *Molecular Plant-Microbe Interactions, 20*(4), 430-440. doi:10.1094/MPMI-20-4-0430

Russell, P. E. (2005). A century of fungicide evolution. *The Journal of Agricultural Science, 143*(1), 11-25. doi:10.1017/S0021859605004971

Sentausa, E., & Fournier, P. E. (2013). Advantages and limitations of genomics in prokaryotic taxonomy. *Clinical Microbiology and Infection, 19*(9), 790-795. doi:10.1111/1469-0691.12181

Sharma, R. R., Singh, D., & Singh, R. (2009). Biological control of postharvest diseases of fruits and vegetables by microbial antagonists: A review. *Biological Control, 50*(3), 205-221. doi:https://doi.org/10.1016/j.biocontrol.2009.05.001

Stein, T., Vater, J., Kruft, V., Otto, A., WittmannLiebold, B., Franke, P., . . . Morris, H. R. (1996). The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. *Journal of Biological Chemistry, 271*(26), 15428-15435.

Steller, S., & Vater, J. (2000). Purification of the fengycin synthetase multienzyme system from *Bacillus subtilis* b213. *Journal of Chromatography B, 737*(1-2), 267-275. doi:10.1016/s0378-4347(99)00481-8

Steller, S., Vollenbroich, D., Leenders, F., Stein, T., Conrad, B., Hofemeister, J., . . . Vater, J. (1999). Structural and functional organization of the fengycin synthetase multienzyme system from *Bacillus subtilis* b213 and A1/3. *Chemistry & Biology, 6*(1), 31-41. doi:10.1016/s1074-5521(99)80018-0

Tseng, C. C., Bruner, S. D., Kohli, R. M., Marahiel, M. A., Walsh, C. T., & Sieber, S. A. (2002). Characterization of the surfactin synthetase C-terminal thioesterase domain as a cyclic depsipeptide synthase. *Biochemistry, 41*(45), 13350-13359. doi:10.1021/bi026592a

Tsuge, K., Akiyama, T., & Shoda, M. (2001). Cloning, sequencing, and characterization of the iturin A operon. *Journal of Bacteriology, 183*(21), 6265-6273. doi:10.1128/jb.183.21.6265-6273.2001

Tsuge, K., Ano, T., Hirai, M., Nakamura, Y., & Shoda, M. (1999). The genes degQ, pps, and lpa-8 (sfp) are responsible for conversion of *Bacillus subtilis* 168 to plipastatin production. *Antimicrobial Agents and Chemotherapy, 43*(9), 2183-2192.

Vining, L. C. (1990). Functions of secondary metabolites. *Annual Review of Microbiology, 44*(1), 395-427. doi:10.1146/annurev.mi.44.100190.002143

Walsh, C. T. (2008). The chemical versatility of natural-product assembly lines. *Accounts of Chemical Research, 41*(1), 4-10. doi:10.1021/ar7000414

Wise, C., Falardeau, J., Hagberg, I., & Avis, T. J. (2014). Cellular lipid composition affects sensitivity of plant pathogens to fengycin, an antifungal compound produced by *Bacillus subtilis* strain CU12. *Phytopathology, 104*(10), 1036-1041. doi:10.1094/phyto-12-13-0336-r

Wise, C., Novitsky, L., Tsopmo, A., & Avis, T. J. (2012). Production and antimicrobial activity of 3-hydroxypropionaldehyde from *Bacillus subtilis* strain CU12. *Journal of Chemical Ecology, 38*(12), 1521-1527. doi:10.1007/s10886-012-0219-2

Yamamoto, S., Kasai, H., Arnold, D. L., Jackson, R. W., Vivian, A., & Harayama, S. (2000). Phylogeny of the genus *Pseudomonas*: intrageneric structure reconstructed from the nucleotide sequences of gyrB and rpoD genes. *Microbiology-Uk, 146*, 2385-2394.

Yang, D. J., Wang, B., Wang, J. X., Chen, Y., & Zhou, M. G. (2009). Activity and efficacy of *Bacillus subtilis* strain NJ-18 against rice sheath blight and Sclerotinia stem rot of rape. *Biological Control, 51*(1), 61-65. doi:10.1016/j.biocontrol.2009.05.021

Yu, X. M., Ai, C. X., Xin, L., & Zhou, G. F. (2011). The siderophore-producing bacterium, *Bacillus subtilis* CAS15, has a biocontrol effect on Fusarium wilt and promotes the growth

of pepper. *European Journal of Soil Biology, 47*(2), 138-145. doi:10.1016/j.ejsobi.2010.11.001

Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research, 18*(5), 821-829. doi:10.1101/gr.074492.107

Zhang, W., Du, P. C., Zheng, H., Yu, W. W., Wan, L., & Chen, C. (2014). Whole-genome sequence comparison as a method for improving bacterial species definition. *Journal of General and Applied Microbiology, 60*(2), 75-78. doi:10.2323/jgam.60.75

Appendixes

Table A1: Nucleotide and protein BLAST results of CLP genes for *Bacillus subtilis* isolate B9-1

Gene	Blastn ID	BLAST Results (protein)	Max score	E value	Blastp ID	GenBank acc. No.
fenA	79%	Nonribosomal peptide synthetase (*Bacillus subtilis*)	2424	0	99%	WP_051484055.1
fenB	81%	Nonribosomal peptide synthase (*Bacillus subtilis*)	2656	0	100%	WP_080015512.1
fenC	79%	Nonribosomal peptide synthase (*Bacillus subtilis*)	2343	0	100%	WP_003231491.1
fenD	81%	Nonribosomal peptide synthetase (*Bacillus subtilis*)	5309	0	99%	WP_069322613.1
fenE	80%	Nonribosomal peptide synthetase (*Bacillus subtilis*)	2473	0	100%	WP_069322612.1
ppsA	98%	plipastatin synthase subunit A (*Bacillus*)	2338	0	100%	WP_010886523.1
ppsB	98%	plipastatin synthase subunit B (*Bacillus subtilis*)	5317	0	99%	WP_019712401.1
ppsC	98%	plipastatin synthase subunit C (*Bacillus*)	2472	0	99%	WP_009967356.1
ppsD	98%	plipastatin synthase subunit D (*Bacillus subtilis*)	2426	0	99%	WP_023592786.1
ppsE	98%	plipastatin synthase subunit E (*Bacillus*)	2648	0	99%	WP_010886522.1
srfAA	99%	surfactin synthase subunit 1 (*Bacillus*)	1221	0	100%	WP_010886402.1
srfAB	100%	surfactin synthase subunit 2 *(Bacillus*)	826	0	99%	WP_010886403.1
srfAC	98%	surfactin synthase subunit 3 (*Bacillus*)	2647	0	100%	WP_003234570.1
srfAD	99%	surfactin synthase thioesterase subunit (*Bacillus subtilis*)	499	3e-179	99%	WP_015715230.1

Table A2: Nucleotide and protein BLAST results of CLP genes for *Bacillus subtilis* isolate B9-5

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No.
fenA	79%	nonribosomal peptide synthetase (*Bacillus subtilis*)	2419	0	99%	WP_048654898.1
fenB	81%	nonribosomal peptide synthetase (*Bacillus subtilis*)	2657	0	100%	WP_015251982.1
fenC	79%	nonribosomal peptide synthetase (*Bacillus subtilis*)	2476	0	98%	WP_003231491.1
fenD	81%	nonribosomal peptide synthetase (*Bacillus subtilis*)	5291	0	99%	WP_038427738.1
fenE	81%	nonribosomal peptide synthetase (*Bacillus subtilis*)	2694	0	99%	WP_044052520.1
ppsA	98%	plipastatin synthase subunit A (*Bacillus*)	2470	0	98%	WP_010886523.1
ppsB	98%	plipastatin synthase subunit B (*Bacillus subtilis*)	5214	0	98%	WP_01971240.1
ppsC	98%	plipastatin synthase subunit C (*Bacillus subtilis*)	2628	0	97%	WP_009968965.1
ppsD	98%	plipastatin synthase subunit D (*Bacillus subtilis*)	2357	0	97%	WP_023592786.1
ppsE	100%	plipastatin synthase subunit E (*Bacillus*)	2614	0	98%	WP_010886522.1
srfAA	99%	surfactin synthase subunit 1 (*Bacillus*)	845	0	99%	WP_010886402.1
srfAB	98%	surfactin synthetase subunit 2 SrfAB (*Bacillus subtilis*)	864	0	99%	WP_080475115.1
srfAC	98%	surfactin synthetase subunit 3 SrfAC (*Bacillus subtilis*)	2528	0	95%	WP_003240142.1
srfAD	98%	surfactin synthase thioesterase subunit (*Bacillus subtilis*)	213	1e-68	100%	WP_015715230.1

Table A3: Nucleotide and protein BLAST results of CLP genes for *Bacillus subtilis* isolate B9-7

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No.
fenA	79%	non-ribosomal peptide synthetase (*Bacillus subtilis*)	2423	0	99%	WP_051484055.1
fenB	81%	non-ribosomal peptide synthetase (*Bacillus subtilis*)	2656	0	100%	WP_080015512.1
fenC	79%	non-ribosomal peptide synthetase (*Bacillus subtilis*)	2518	0	99%	WP_032730646.1
fenD	81%	non-ribosomal peptide synthetase (*Bacillus subtilis*)	5309	0	99%	WP_069322613.1
fenE	80%	non-ribosomal peptide synthetase (*Bacillus subtilis*)	2473	0	100%	WP_069322612.1
ppsA	98%	plipastatin synthase subunit A (*Bacillus*)	2516	0	99%	WP_010886523.1
ppsB	98%	plipastatin synthase subunit B (*Bacillus subtilis*)	5317	0	99%	WP_019712401.1
ppsC	98%	plipastatin synthase subunit C (*Bacillus subtilis*)	2471	0	99%	WP_009968965.1
ppsD	98%	plipastatin synthase subunit D (*Bacillus subtilis*)	2426	0	99%	WP_023592786.1
ppsE	98%	plipastatin synthase subunit E (*Bacillus*)	2648	0	99%	WP_010886522.1
srfAA	100%	surfactin synthase subunit 1 (*Bacillus*)	846	0	100%	WP_010886402.1
srfAB	100%	surfactin synthetase subunit 2 SrfAB (*Bacillus subtilis)*	821	0	99%	WP_080360745.1
srfAC	98%	surfactin synthetase subunit 3 SrfAC (*Bacillus subtilis*)	2529	0	95%	WP_003240142.1
srfAD	99%	surfactin synthase thioesterase subunit (*Bacillus subtilis*)	495	8e-178	99%	WP_015715230.1

Table A4: Nucleotide and protein BLAST results of CLP genes for *Bacillus subtilis* isolate B9-8

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No.
fenA	79%	non-ribosomal peptide synthetase (*Bacillus subtilis*)	2419	0	99%	WP_048654898.1
fenB	81%	non-ribosomal peptide synthetase (*Bacillus subtilis*)	2657	0	100%	WP_015251982.1
fenC	79%	non-ribosomal peptide synthetase (*Bacillus subtilis*)	2476	0	98%	WP_003231491.1
fenD	81%	non-ribosomal peptide synthetase (*Bacillus subtilis*)	5291	0	99%	WP_038427738.1
fenE	81%	non-ribosomal peptide synthetase (*Bacillus subtilis*)	2694	0	99%	WP_044052520.1
ppsA	98%	plipastatin synthase subunit A (*Bacillus*)	2470	0	98%	WP_010886523.1
ppsB	98%	plipastatin synthase subunit B (*Bacillus subtilis*)	5214	0	98%	WP_019712401.1
ppsC	98%	plipastatin synthase subunit C (*Bacillus subtilis*)	2628	0	97%	WP_009968965.1
ppsD	98%	plipastatin synthase subunit D (*Bacillus subtilis*)	2357	0	97%	WP_023592786.1
ppsE	98%	plipastatin synthase subunit E *(Bacillus*)	2614	0	98%	WP_010886522.1
srfAA	99%	surfactin synthase subunit 1 (*Bacillus*)	845	0	99%	WP_010886402.1
srfAB	98%	surfactin synthetase subunit 2 SrfAB (*Bacillus subtilis*)	864	0	99%	WP_080475115.1
srfAC	98%	surfactin synthase subunit 3 (*Bacillus*)	2618	0	99%	WP_003234570.1
srfAD	98%	surfactin synthase thioesterase subunit (*Bacillus subtilis*)	213	1e-68	100%	WP_015715230.1

Table A5: Nucleotide and protein BLAST results of CLP genes for *Bacillus subtilis* isolate B9-9A

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No.
fenA	79%	nonribosomal peptide synthase (*Bacillus subtilis*)	2455	0	99%	WP_051484055.1
fenB	81%	nonribosomal peptide synthase (*Bacillus subtilis*)	2656	0	100%	WP_080015512.1
fenC	79%	nonribosomal peptide synthase (*Bacillus subtilis*)	2343	0	100%	WP_003231491.1
fenD	81%	nonribosomal peptide synthase *(Bacillus subtilis*)	5309	0	99%	WP_069322613.1
fenE	80%	nonribosomal peptide synthase (*Bacillus subtilis*)	2406	0	99%	WP_019846805.1
ppsA	98%	plipastatin synthase subunit A (*Bacillus*)	2338	0	100%	WP_010886523.1
ppsB	98%	plipastatin synthase subunit B *(Bacillus subtilis*)	5317	0	99%	WP_019712401.1
ppsC	98%	plipastatin synthase subunit C (*Bacillus subtilis*)	2403	0	99%	WP_009968965.1
ppsD	98%	plipastatin synthase subunit D (*Bacillus subtilis*)	2458	0	99%	WP_023592786.1
ppsE	98%	plipastatin synthase subunit E (*Bacillus*)	2648	0	99%	WP_010886522.1
srfAA	100%	surfactin synthase subunit 1 (*Bacillus*)	846	0	100%	WP_010886402.1
srfAB	100%	surfactin synthetase subunit 2 SrfAB (*Bacillus subtilis*)	840	0	99%	WP_080333204.1
srfAC	98%	surfactin synthetase subunit 3 SrfAC (*Bacillus subtilis*)	2529	0	95%	WP_003240142.1
srfAD	99%	surfactin synthase thioesterase subunit (*Bacillus subtilis*)	495	7e-178	99%	WP_015715230.1

Table A6: Nucleotide and protein BLAST results of CLP genes for *Bacillus megaterium* isolate B9-9B

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No
fenA	78%	2-succinylbenzoate-CoA ligase (*Bacillus megaterium*)	1001	0	99%	WP_075419767.1
fenB	-	-	-	-	-	-
fenC	-	-	-	-	-	-
fenD	85%	acyl-CoA ligase (*Bacillus megaterium*)	1081	0	99%	WP_047932434.1
fenE	-	-	-	-	-	-
ppsA	-	-	-	-	-	-
ppsB	-	-	-	-	-	-
ppsC	-	-	-	-	-	-
ppsD	78%	2-succinylbenzoate-CoA ligase (*Bacillus megaterium*)	1001	0	99%	WP_075419767.1
ppsE	-	-	-	-	-	-
srfAA	74%	pyroglutamyl-peptidase I (*Bacillus megaterium*)	422	2e-149	97%	WP_034653849.1
srfAB	84%	long-chain fatty acid-CoA ligase (*Bacillus megaterium*)	1046	0	99%	WP_074897532.1
srfAC	-	-	-	-	-	-
srfAD	-	-	-	-	-	-

Table A7: Nucleotide and protein BLAST results of CLP genes for *Bacillus subtilis* isolate B9-14

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No.
fenA	79%	non-ribosomal peptide synthase (*Bacillus subtilis*)	2423	0	99%	WP_051484055.1
fenB	81%	non-ribosomal peptide synthase (*Bacillus subtilis*)	2649	0	99%	WP_080031204.1
fenC	79%	non-ribosomal peptide synthase (*Bacillus subtilis*)	2342	0	100%	WP_003247157.1
fenD	81%	non-ribosomal peptide synthase (*Bacillus subtilis*)	5309	0	99%	WP_069322613.1
fenE	80%	non-ribosomal peptide synthase (*Bacillus subtilis*)	2469	0	99%	WP_032730497.1
ppsA	98%	plipastatin synthase subunit A (*Bacillus*)	2338	0	100%	WP_010886523.1
ppsB	98%	plipastatin synthase subunit B (*Bacillus subtilis*)	5317	0	99%	WP_019712401.1
ppsC	98%	plipastatin synthase subunit C (*Bacillus subtilis*)	2471	0	99%	WP_009968965.1
ppsD	98%	plipastatin synthase subunit D (*Bacillus subtilis*)	2426	0	99%	WP_023592786.1
ppsE	98%	plipastatin synthase subunit E (*Bacillus*)	2648	0	99%	WP_010886522.1
srfAA	100%	surfactin synthase subunit 1 (*Bacillus*)	846	0	100%	WP_010886402.1
srfAB	100%	surfactin synthetase subunit 2 SrfAB (*Bacillus subtilis*)	840	0	99%	WP_080333204.1
srfAC	98%	surfactin synthetase subunit 3 SrfAC (*Bacillus subtilis*)	2529	0	95%	WP_003240142.1
srfAD	99%	surfactin synthase thioesterase subunit (*Bacillus subtilis*)	499	3e-179	99%	WP_015715230.1

Table A8: Nucleotide and protein BLAST results of CLP genes for *Bacillus subtilis* isolate CU12

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No.
fenA	80%	fengycin synthetase FenA (*Bacillus subtilis*)	5955	0	80%	AAB80955.2
fenB	81%	nonribosomal peptide synthetase (*Bacillus subtilis*)	2637	0	99%	WP_088467231.1
fenC	79%	nonribosomal peptide synthetase (*Bacillus subtilis*)	2340	0	100%	WP_068947606.1
fenD	81%	nonribosomal peptide synthetase (*Bacillus subtilis*)	5322	0	99%	WP_038427738.1
fenE	80%	nonribosomal peptide synthetase (*Bacillus subtilis*)	2423	0	99%	WP_068947605.1
ppsA	97%	plipastatin synthase subunit A (*Bacillus subtilis*)	2296	0	98%	WP_010886523.1
ppsB	98%	plipastatin synthase subunit B (*Bacillus subtilis*)	5209	0	98%	WP_019712401.1
ppsC	98%	plipastatin synthase subunit C (*Bacillus subtilis*)	2386	0	98%	CUB50477.1
ppsD	98%	plipastatin synthase subunit D (*Bacillus*)	7344	0	98%	WP_009967354.1
ppsE	98%	plipastatin synthase subunit E (*Bacillus subtilis*)	2625	0	99%	AIX07617.1
srfAA	98%	surfactin synthase subunit 1 (*Bacillus*)	1450	0	99%	WP_010886402.1
srfAB	98%	surfactin synthetase subunit 2 SrfAB (*Bacillus subtilis*)	835	0	98%	WP_080475115.1
srfAC	99%	surfactin synthetase subunit 3 SrfAC (*Bacillus subtilis*)	2534	0	95%	WP_003240142.1
srfAD	99%	surfactin synthase thioesterase subunit (*Bacillus subtilis*)	499	5e-179	99%	WP_015715230.1
ituD	70%	[acyl-carrier-protein] S-malonyltransferase (*Bacillus subtilis*)	631	0	99%	WP_015383644.1
ituA	69%	nonribosomal peptide synthetase (*Bacillus subtilis*)	7460	0	99%	WP_038427737.1
ituB	72%	nonribosomal peptide synthetase (*Bacillus subtilis*)	2423	0	99%	WP_068947605.1
ituC	68%	nonribosomal peptide synthetase (*Bacillus subtilis*)	2397	0	99%	WP_088326144.1

Table A9: Nucleotide and protein BLAST results of CLP genes for *Bacillus subtilis* isolate F9-2

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No.
fenA	80%	nonribosomal peptide synthetase (*Bacillus subtilis*)	3659	0	99%	WP_051484055.1
fenB	81%	nonribosomal peptide synthetase (*Bacillus subtilis*)	2656	0	100%	WP_080015512.1
fenC	80%	nonribosomal peptide synthetase (*Bacillus subtilis*)	5293	0	99%	WP_019846737.1
fenD	81%	nonribosomal peptide synthetase (*Bacillus subtilis*)	5309	0	99%	WP_069322613.1
fenE	82%	nonribosomal peptide synthetase (*Bacillus subtilis*)	4408	0	85%	WP_019846737.1
ppsA	98%	plipastatin synthase subunit A (*Bacillus*)	5261	0	100%	WP_010886523.1
ppsB	98%	plipastatin synthase subunit B (*Bacillus subtilis*)	5317	0	99%	WP_019712401.1
ppsC	96%	plipastatin synthase subunit C (*Bacillus subtilis*)	3699	0	72%	WP_009968965.1
ppsD	98%	plipastatin synthase subunit D (*Bacillus subtilis*)	3713	0	99%	WP_023592786.1
ppsE	98%	plipastatin synthase subunit E (*Bacillus subtilis*)	2622	0	99%	AIX07617.1
srfAA	99%	surfactin synthase subunit 1 (*Bacillus*)	2507	0	99%	WP_046380780.1
srfAB	99%	surfactin synthetase subunit 2 SrfAB (*Bacillus subtilis*)	840	0	99%	WP_080333204.1
srfAC	100%	surfactin synthetase subunit 3 SrfAC (*Bacillus subtilis*)	2529	0	95%	WP_003240142.1
srfAD	99%	surfactin synthase thioesterase subunit (*Bacillus subtilis*)	499	5e-179	99%	WP_015715230.1

Table A10: Nucleotide and protein BLAST results of CLP genes for *Bacillus subtilis* isolate F9-8

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No
fenA	79%	nonribosomal peptide synthase (*Bacillus subtilis*)	2383	0	98%	WP_031600579.1
fenB	81%	nonribosomal peptide synthase (*Bacillus subtilis*)	2656	0	100%	WP_080015512.1
fenC	79%	nonribosomal peptide synthase (*Bacillus subtilis*)	2343	0	100%	WP_003231491.1
fenD	81%	nonribosomal peptide synthase (*Bacillus subtilis*)	5309	0	99%	WP_069322613.1
fenE	80%	nonribosomal peptide synthase (*Bacillus subtilis*)	2473	0	100%	WP_069322612.1
ppsA	98%	plipastatin synthase subunit A (*Bacillus*)	2338	0	100%	WP_010886523.1
ppsB	98%	plipastatin synthase subunit B (*Bacillus subtilis*)	5317	0	99%	WP_019712401.1
ppsC	98%	plipastatin synthase subunit C (*Bacillus subtilis*)	2471	0	99%	WP_009968965.1
ppsD	98%	plipastatin synthase subunit D (*Bacillus*)	2424	0	99%	WP_009967354.1
ppsE	98%	plipastatin synthase subunit E (*Bacillus*)	2648	0	99%	WP_010886522.1
srfAA	100%	surfactin synthase subunit 1 (*Bacillus*)	798	0	99%	WP_010886402.1
srfAB	100%	surfactin synthetase subunit 2 SrfAB (*Bacillus subtilis*)	821	0	99%	WP_080360745.1
srfAC	98%	surfactin synthetase subunit 3 SrfAC (*Bacillus subtilis*)	2529	0	95%	WP_003240142.1
srfAD	99%	surfactin synthase thioesterase subunit (*Bacillus subtilis*)	499	3e-179	99%	WP_015715230.1
ituD	70%	[acyl-carrier-protein] S-malonyltransferase (*Bacillus subtilis*)	634	0	99%	WP_064671373.1
ituA	70%	nonribosomal peptide synthetase (*Bacillus subtilis*)	11407	0	100%	WP_080276611.1
ituB	80%	nonribosomal peptide synthetase (*Bacillus subtilis*)	5309	0	99%	WP_069322613.1
ituC	80%	nonribosomal peptide synthetase (*Bacillus subtilis*)	2338	0	100%	WP_009967358.1

Table A11: Nucleotide and protein BLAST results of CLP genes for *Bacillus subtilis* isolate F9-12

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No
fenA	79%	nonribosomal peptide synthetase (*Bacillus subtilis*)	2455	0	99%	WP_051484055.1
fenB	81%	nonribosomal peptide synthase (*Bacillus subtilis*)	2656	0	100%	WP_080015512.1
fenC	81%	nonribosomal peptide synthase *(Bacillus subtilis*)	2343	0	100%	WP_003231491.1
fenD	81%	nonribosomal peptide synthetase (*Bacillus subtilis*)	5309	0	99%	WP_069322613.1
fenE	81%	nonribosomal peptide synthetase (*Bacillus subtilis*)	3683	0	99%	WP_048217643.1
ppsA	98%	plipastatin synthase subunit A (*Bacillus*)	2338	0	100%	WP_010886523.1
ppsB	98%	plipastatin synthase subunit B (*Bacillus subtilis*)	5317	0	99%	WP_019712401.1
ppsC	98%	plipastatin synthase subunit C (*Bacillus subtilis*)	3682	0	99%	WP_009968965.1
ppsD	98%	plipastatin synthase subunit D (*Bacillus subtilis*)	2458	0	99%	WP_023592786.1
ppsE	98%	plipastatin synthase subunit E (*Bacillus*)	2648	0	99%	WP_010886522.1
srfAA	99%	surfactin synthase subunit 1 (*Bacillus*)	1628	0	99%	WP_010886402.1
srfAB	100%	surfactin synthetase subunit 2 SrfAB (*Bacillus subtilis*)	846	0	100%	WP_080360745.1
srfAC	98%	surfactin synthetase subunit 3 SrfAC (*Bacillus subtilis*)	2529	0	95%	WP_003240142.1
srfAD	99%	surfactin synthase thioesterase subunit (*Bacillus subtilis*)	499	3e-179	99%	WP_015715230.1

Table A12: Nucleotide and protein BLAST results of CLP genes for *Bacillus megaterium* isolate M9-1B

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No
fenA	78%	2-succinylbenzoate-CoA ligase (*Bacillus megaterium*)	1001	0	99%	WP_075419767.1
fenB	-	-	-	-	-	-
fenC	-	-	-	-	-	-
fenD	85%	acyl-CoA ligase (*Bacillus megaterium*)	1081	0	99%	WP_047932434.1
fenE	-	-	-	-	-	-
ppsA	-	-	-	-	-	-
ppsB	-	-	-	-	-	-
ppsC	-	-	-	-	-	-
ppsD	78%	2-succinylbenzoate-CoA ligase (*Bacillus megaterium*)	1001	0	99%	WP_075419767.1
ppsE	-	-	-	-	-	-
srfAA	74%	pyroglutamyl-peptidase I (*Bacillus megaterium*)	422	2e-149	97%	WP_034653849.1
srfAB	84%	long-chain fatty acid-CoA ligase (*Bacillus megaterium*)	1046	0	99%	WP_074897532.1
srfAC	-	-	-	-	-	-
srfAD	-	-	-	-	-	-

Table A13: Nucleotide and protein BLAST results of CLP genes for *Bacillus subtilis* isolate M9-3

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No
fenA	79%	nonribosomal peptide synthetase (*Bacillus subtilis*)	2399	0	98%	WP_080481276.1
fenB	81%	nonribosomal peptide synthase (*Bacillus subtilis*)	2655	0	100%	WP_077671723.1
fenC	79%	nonribosomal peptide synthase (*Bacillus subtilis*)	2324	0	99%	WP_086343906.1
fenD	81%	nonribosomal peptide synthetase (*Bacillus subtilis*)	5267	0	99%	WP_032677285.1
fenE	81%	nonribosomal peptide synthetase (*Bacillus subtilis*)	3479	0	98%	WP_047182691.1
ppsA	99%	plipastatin synthase subunit A (*Bacillus*)	2294	0	98%	WP_010886523.1
ppsB	98%	plipastatin synthase subunit B (*Bacillus subtilis*)	5226	0	98%	WP_019712401.1
ppsC	98%	plipastatin synthase subunit C (*Bacillus subtilis*)	3432	0	97%	WP_009968965.1
ppsD	97%	plipastatin synthase subunit D (*Bacillus subtilis*)	2366	0	97%	WP_023592786.1
ppsE	98%	plipastatin synthase subunit E (*Bacillus*)	2579	0	97%	WP_010886522.1
srfAA	98%	surfactin synthase subunit 1 (*Bacillus*)	835	0	99%	WP_010886402.1
srfAB	97%	surfactin synthetase subunit 2 SrfAB (*Bacillus subtilis*)	836	0	98%	WP_080326372.1
srfAC	98%	surfactin synthetase subunit 3 SrfAC (*Bacillus subtilis*)	2523	0	95%	WP_003240142.1
srfAD	99%	surfactin synthase thioesterase subunit (*Bacillus subtilis*)	498	6e-179	100%	WP_04297202.1

Table A14: Nucleotide and protein BLAST results of CLP genes for *Bacillus subtilis* isolate M9-4

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No
fenA	81%	nonribosomal peptide synthetase (*Bacillus subtilis*)	7461	0	99%	WP_051484055.1
fenB	81%	nonribosomal peptide synthase (*Bacillus subtilis*)	2656	0	100%	WP_080015512.1
fenC	79%	nonribosomal peptide synthase *(Bacillus subtilis*)	2343	0	100%	WP_003231491.1
fenD	80%	nonribosomal peptide synthetase (*Bacillus subtilis*)	5845	0	80%	WP_050820704.1
fenE	80%	nonribosomal peptide synthase (*Bacillus subtilis*)	2424	0	99%	WP_019846805.1
ppsA	98%	plipastatin synthase subunit A (*Bacillus*)	2338	0	100%	WP_010886523.1
ppsB	97%	nonribosomal peptide synthetase (*Bacillus subtilis*)	5845	0	80%	WP_050820704.1
ppsC	98%	plipastatin synthase subunit C (*Bacillus subtilis*)	2422	0	99%	WP_009968965.1
ppsD	98%	plipastatin synthase subunit D (*Bacillus*)	7433	0	99%	WP_009967354.1
ppsE	98%	plipastatin synthase subunit E (*Bacillus*)	2648	0	99%	WP_010886522.1
srfAA	99%	surfactin synthase subunit 1 (*Bacillus*)	5845	0	80%	WP_010886402.1
srfAB	100%	surfactin synthetase subunit 2 SrfAB (*Bacillus subtilis*)	846	0	100%	WP_080360745.1
srfAC	98%	surfactin synthetase subunit 3 SrfAC (*Bacillus subtilis*)	2529	0	95%	WP_003240142.1
srfAD	99%	surfactin synthase thioesterase subunit (*Bacillus subtilis*)	499	3e-179	99%	WP_015715230.1

Table A15: Nucleotide and protein BLAST results of CLP genes for *Bacillus subtilis* isolate M9-7

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No
fenA	81%	nonribosomal peptide synthetase (*Bacillus subtilis*)	7465	0	99%	WP_051484055.1
fenB	81%	nonribosomal peptide synthase (*Bacillus subtilis*)	2656	0	100%	WP_080015512.1
fenC	79%	nonribosomal peptide synthase (*Bacillus subtilis*)	2343	0	100%	WP_003231491.1
fenD	80%	nonribosomal peptide synthetase (*Bacillus subtilis*)	2440	0	100%	WP_032723105.1
fenE	80%	nonribosomal peptide synthetase (*Bacillus subtilis*)	2473	0	100%	WP_069322612.1
ppsA	98%	plipastatin synthase subunit A (*Bacillus*)	2338	0	100%	WP_010886523.1
ppsB	97%	plipastatin synthase subunit B (*Bacillus subtilis*)	2439	0	100%	WP_019712401.1
ppsC	98%	plipastatin synthase subunit C (*Bacillus subtilis*)	2471	0	99%	WP_009968965.1
ppsD	98%	plipastatin synthase subunit D (*Bacillus*)	7434	0	99%	WP_009967354.1
ppsE	98%	plipastatin synthase subunit E (*Bacillus*)	2648	0	99%	WP_010886522.1
srfAA	100%	surfactin synthase subunit 1 (*Bacillus*)	798	0	99%	WP_010886402.1
srfAB	100%	surfactin synthetase subunit 2 SrfAB (*Bacillus subtilis*)	821	0	99%	WP_080360745.1
srfAC	98%	surfactin synthetase subunit 3 SrfAC (*Bacillus subtilis*)	2529	0	95%	WP_003240142.1
srfAD	99%	surfactin synthase thioesterase subunit (*Bacillus subtilis*)	499	3e-179	99%	WP_015715230.1

Table A16: Nucleotide and protein BLAST results of CLP genes for *Bacillus subtilis* isolate M9-9

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No
fenA	79%	nonribosomal peptide synthetase *(Bacillus subtilis*)	2423	0	99%	WP_051481055.1
fenB	81%	nonribosomal peptide synthase (*Bacillus subtilis*)	2656	0	100%	WP_080015512.1
fenC	79%	nonribosomal peptide synthase (*Bacillus subtilis*)	2343	0	100%	WP_0032301491.1
fenD	81%	nonribosomal peptide synthetase (*Bacillus subtilis*)	5309	0	99%	WP_069322613.1
fenE	80%	nonribosomal peptide synthase (*Bacillus subtilis*)	2473	0	99%	WP_019846805.1
ppsA	98%	plipastatin synthase subunit A (*Bacillus*)	2338	0	100%	WP_010886523.1
ppsB	98%	plipastatin synthase subunit B (*Bacillus subtilis*)	5317	0	99%	WP_019712401.1
ppsC	98%	plipastatin synthase subunit C *(Bacillus subtilis*)	2471	0	99%	WP_009968965.1
ppsD	98%	plipastatin synthase subunit D (*Bacillus subtilis*)	2426	0	99%	WP_023592786.1
ppsE	98%	plipastatin synthase subunit E (*Bacillus*)	2648	0	99%	WP_010886522.1
srfAA	100%	surfactin synthase subunit 1 (*Bacillus*)	846	0	100%	WP_010886402.1
srfAB	100%	surfactin synthetase subunit 2 SrfAB (*Bacillus subtilis*)	846	0	100%	WP_080360745.1
srfAC	98%	surfactin synthetase subunit 3 SrfAC (*Bacillus subtilis*)	2529	0	95%	WP_003240142.1
srfAD	99%	surfactin synthase thioesterase subunit *(Bacillus subtilis*)	499	3e-179	99%	WP_015715230.1
ituD	70%	[acyl-carrier-protein] S-malonyltransferase (*Bacillus subtilis*)	634	0	99%	WP_064671373.1
ituA	70%	nonribosomal peptide synthetase (*Bacillus subtilis*)	11328	0	99%	WP_080287613.1
ituB	80%	nonribosomal peptide synthetase (*Bacillus subtilis*)	520	0	98%	WP_072175171.1
ituC	71%	nonribosomal peptide synthetase (*Bacillus subtilis*)	2396	0	99%	WP_074794567.1

Table A17: Nucleotide and protein BLAST results of CLP genes for *Bacillus subtilis* isolate M9-14

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No
fenA	79%	nonribosomal peptide synthetase (*Bacillus subtilis*)	2455	0	99%	WP_051484055.1
fenB	81%	nonribosomal peptide synthase (*Bacillus subtilis*)	2656	0	100%	WP_080015512.1
fenC	80%	nonribosomal peptide synthetase (*Bacillus subtilis*)	4405	0	85%	WP_019846737.1
fenD	80%	nonribosomal peptide synthetase (*Bacillus subtilis*)	5845	0	80%	WP_019846924.1
fenE	80%	nonribosomal peptide synthase (*Bacillus subtilis*)	2473	0	99%	WP_019846805.1
ppsA	98%	plipastatin synthase subunit A (*Bacillus*)	4376	0	84%	WP_010886523.1
ppsB	97%	nonribosomal peptide synthetase (*Bacillus subtilis*)	5845	0	80%	WP_019846924.1
ppsC	96%	plipastatin synthase subunit C *(Bacillus subtilis*)	2471	0	99%	WP_009968965.1
ppsD	98%	plipastatin synthase subunit D (*Bacillus subtilis*)	2458	0	99%	WP_02359786.1
ppsE	98%	plipastatin synthase subunit E (*Bacillus*)	2648	0	99%	WP_010886522.1
srfAA	99%	surfactin synthase subunit 1 (*Bacillus*)	7394	0	99%	WP_010886402.1
srfAB	99%	surfactin synthase subunit 2 (*Bacillus*)	7401	0	99%	WP_010886403.1
srfAC	98%	surfactin synthetase subunit 3 SrfAC (*Bacillus subtilis*)	2529	0	95%	WP_003240142.1
srfAD	99%	surfactin synthase thioesterase subunit (*Bacillus subtilis*)	499	3e-179	99%	WP_015715230.1

Table A18: Nucleotide and protein BLAST results of CLP genes for *Brevibacillus borstelensis* isolate M9-18

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No.
fenA	67%	nonribosomal peptide synthetase (*Brevibacillus borstelensis*)	7114	0	99%	WP_003389551.1
fenB	67%	nonribosomal peptide synthetase (*Brevibacillus borstelensis*)	5649	0	99%	WP_031932063.1
fenC	66%	nonribosomal peptide synthetase (*Brevibacillus borstelensis*)	5900	0	99%	WP_003389552.1
fenD	77%	nonribosomal peptide synthetase (*Brevibacillus borstelensis*)	5895	0	99%	WP_024983402.1
fenE	76%	nonribosomal peptide synthetase (*Brevibacillus borstelensis*)	5646	0	99%	WP_003389361.1
ppsA	68%	nonribosomal peptide synthetase (*Brevibacillus borstelensis*)	5900	0	99%	WP_003389552.1
ppsB	75%	nonribosomal peptide synthetase *(Brevibacillus borstelensis*)	5895	0	99%	WP_024983402.1
ppsC	70%	nonribosomal peptide synthetase *(Brevibacillus borstelensis*)	5646	0	99%	WP_003389361.1
ppsD	69%	nonribosomal peptide synthetase (*Brevibacillus borstelensis*)	7114	0	99%	WP_003389551.1
ppsE	72%	nonribosomal peptide synthetase (*Brevibacillus borstelensis*)	5649	0	99%	WP_031932063.1
srfAA	75%	nonribosomal peptide synthetase (*Brevibacillus borstelensis*)	7158	0	99%	WP_051925351.1
srfAB	86%	nonribosomal peptide synthetase (*Brevibacillus borstelensis*)	5902	0	99%	WP_030191899.1
srfAC	77%	nonribosomal peptide synthetase (*Brevibacillus borstelensis*)	7148	0	99%	WP_024983401.1
srfAD	-	-	-	-	-	-

Table A19: Nucleotide and protein BLAST results of CLP genes for *Bacillus badius* isolate M9-20

Gene	Blastn ID	BLAST Results (protein)	Max Score	E value	Blastp ID	GenBank acc. No.
fenA	82%	2-succinylbenzoate-CoA ligase (*Bacillus* sp. FJAT-27231)	867	0	85%	WP_049663911.1
fenB	79%	o-succinylbenzoate-CoA ligase (*Bacillus* sp. FJAT-27231)	865	0	84%	WP_065408713.1
fenC	84%	2-succinylbenzoate-CoA ligase (*Bacillus badius*)	841	0	82%	WP_041114035.1
fenD	71%	o-succinylbenzoate-CoA ligase (*Bacillus* sp. FJAT-27231)	839	0	82%	WP_063441137.1
fenE	84%	2-succinylbenzoate-CoA ligase (*Bacillus badius*)	837	0	82%	WP_041097891.1
ppsA	81%	2-succinylbenzoate-CoA ligase (*Bacillus badius*)	841	0	82%	WP_041114035.1
ppsB	-	-	-	-	-	-
ppsC	81%	2-succinylbenzoate-CoA ligase (*Bacillus badius*)	837	0	82%	WP_041097891.1
ppsD	85%	2-succinylbenzoate-CoA ligase (*Bacillus* sp. FJAT-27231)	867	0	85%	WP_049663911.1
ppsE	80%	o-succinylbenzoate-CoA ligase (*Bacillus* sp. FJAT-27231)	865	0	84%	WP_065408713.1
srfAA	84%	nonribosomal peptide synthetase (*Bacillus badius*)	2085	0	79%	WP_063440830.1
srfAB	84%	long-chain fatty acid-CoA ligase (*Bacillus badius*)	920	0	84%	WP_063441357.1
srfAC	79%	2-succinylbenzoate-CoA ligase (*Bacillus badius*)	841	0	82%	WP_041114035.1
srfAD	-	-	-	-	-	-